

NONPARAMETRIC BERNSTEIN-VON MISES PHENOMENON: A TUNING PRIOR PERSPECTIVE

BY ZUOFENG SHANG^{*} AND GUANG CHENG[†]

Department of Statistics, Purdue University

JULY 17, 2015

Statistical inference on infinite-dimensional parameters in a Bayesian framework is investigated. The main contribution of this paper is to demonstrate that nonparametric Bernstein-von Mises theorem can be established in a *general* class of nonparametric regression models under a novel tuning prior (indexed by a non-random hyper-parameter). Surprisingly, this type of prior connects two important classes of statistical methods: nonparametric Bayes and smoothing spline. The intrinsic connection with smoothing spline greatly facilitates both theoretical analysis and practical implementation for nonparametric Bayesian inference. For example, we can employ generalized cross validation to select a proper tuning prior, under which the constructed credible regions/intervals are frequentist valid. This new methodology is supported by our simulations. The posterior contraction rate (under Sobolev norm) is derived as a by-product. A collection of technical tools such as RKHS theory, Cameron-Martin theorem [6] and Gaussian correlation inequality [32] are employed in this paper.

1. Introduction. A common practice in quantifying Bayesian uncertainty is to construct credible regions (or sets) that cover a large fraction of posterior mass. The construction is built upon posterior distribution whose asymptotic behavior is characterized by Bernstein-von Mises (BvM) theorem. As a consequence of BvM theorem, Bayesian credible sets may asymptotically possess frequentist validity, which is called as Bernstein-von Mises phenomenon in [31]. In nonparametric settings, a BvM result was initially found *impossible* by [12, 17]. An essential reason for this failure is due to the lack of flexibility in the assigned priors. In the paper, we introduce a novel *tuning prior* under which BvM phenomenon can be established in a general class of nonparametric regression models. More interestingly, this type of prior connects two important classes of statistical methods: nonparametric Bayes and smoothing spline. This coupling effect is practically

^{*}Visiting Assistant Professor

[†]Corresponding Author. Associate Professor, Research Sponsored by NSF CAREER Award DMS-1151692, DMS-1418042, Simons Fellowship in Mathematics, Office of Naval Research (ONR N00014-15-1-2331) and a grant from Indiana Clinical and Translational Sciences Institute. Guang Cheng is on sabbatical at Princeton while this work is revised; he would like to thank the Princeton ORFE department for its hospitality and support.

AMS 2000 subject classifications: Primary 62C10 Secondary 62G15, 62G08

Keywords and phrases: Bayesian inference, nonparametric Bernstein-von Mises Theorem, small ball probability, smoothing spline, tuning prior

useful. For example, a proper tuning prior can be selected through generalized cross validation (GCV).

As far as we are aware, BvM results are only shown to be valid in various special settings. For instance, in Gaussian sequence models (equivalently, Gaussian white noise models), BvM theorem has been established for mean sequences (or signals); see [31, 27, 28, 47, 48, 9, 10, 35, 34]. In Gaussian regression with fixed design, [44, 45] proposed adaptive credible regions for regression functions; with random design, [58] proposed credible sets under sieved priors. In models where efficient estimation (at \sqrt{n} -rate) is possible, [39, 9, 10, 11] proposed credible intervals for functionals of an infinite-dimensional parameter. To the best of our knowledge, establishment of BvM theorem (together with BvM phenomenon) in a more general framework remains an open problem. Our major goal in this paper is to develop a general BvM framework that deals with both Gaussian and non-Gaussian data in a unified manner. Moreover, our results allow random design, and are valid even when efficient estimation is unavailable.

We propose a type of tuning prior whose distribution is governed by a (non-random) hyperparameter. In particular, the proposed prior controls the magnitude of higher-order derivatives of regression functions, and therefore, has a natural association with the roughness penalty in smoothing spline. In fact, their precise relation can be discovered by an application of Hájek's Lemma ([21]). An interesting finding is that the tuning prior is induced by a particular type of Gaussian process (GP) (different from the sieved prior considered by [58]) through a Radon-Nikodym derivative. In the special Gaussian regression, these Gaussian processes match with the sequence priors considered in [56]. Thus, our tuning prior can be viewed as an extension of [56] to the more general Bayesian nonparametric framework.

The above discussions imply duality relation between the resulting posterior distribution and the penalized likelihood in the smoothing spline literature. Therefore, we are able to establish BvM results in a nonparametric exponential family by borrowing recent smoothing spline inference results ([41]). This also justifies the use of GCV method in selecting a proper tuning of prior. We want to remark that our new methodology is different from empirical Bayes approach, and supported by our simulation results.

Our BvM theorem holds in total variation distance, extending that obtained by [31] in Gaussian white noise. As a consequence, we construct *exact* credible regions for the regression function under both strong and weak topology, and *exact* credible intervals for its linear functionals, e.g., its local value. Furthermore, the frequentist behaviors of these regions/intervals are carefully investigated. For instance, when the tuning parameter is carefully chosen, the frequentist coverage of the credible set approaches one given any credibility level, called as weak BvM phenomenon. The overly conservativeness of credible sets can be resolved by invoking a weaker topology (inspired by [9, 10]). This leads to the so-called strong BvM phenomenon. In contrast, the BvM theorem developed by [9, 10] is in terms of a weaker metric, i.e., bounded Lipschitz metric, and crucially relies on the *efficiency* of nonparametric estimate ([4]). Other relevant BvM theorems include [13, 14, 7, 3, 39, 11, 22, 25, 26] for semi/nonparametric models or functionals of the density.

An intermediate step in developing our BvM theorem is to derive posterior contraction rates under a type of Sobolev norm, which can be counted as another contribution. This Sobolev norm is stronger than the commonly used metrics in deriving contraction rates such as Kullback-Leibler divergence, Hellinger distance and supremum norm ([19, 20, 51, 52, 53, 54, 18, 8]). The reproducing kernel Hilbert space (RKHS) theory, uniformly consistent test arguments ([20]) together with Cameron-Martin theorem [6] and Gaussian correlation inequality [32] are applied to derive this new result. A relevant result is [58] who derive contraction rates for higher-order derivatives in Gaussian regression. It is worth mentioning that our contraction rate result does not rely on posterior consistency or posterior boundedness in contrast with [18].

The rest of this article is organized as follows. In Section 2, we present a general nonparametric exponential framework including Gaussian regression and logistic regression, and Section 3 constructs the tuning prior that ties nonparametric Bayes models and smoothing spline models. Section 4 includes results on posterior contraction rates under a type of Sobolev norm. Section 5 includes the main result of the article: nonparametric BvM theorem. Section 6 discusses two important applications of BvM theorem: construction of credible region of the regression function and credible interval of a general class of linear functionals. Frequentist validity is also investigated. Section 7 includes a simulation study. Technical details are postponed to either Appendix or a Supplementary Document [42].

2. Nonparametric Regression Model. In this section, we present a general class of nonparametric regression models: nonparametric exponential family. Let $Y \in \mathcal{Y} \subseteq \mathbb{R}$ be response variable and $X \in \mathbb{I} := [0, 1]$ be covariate variable. Our general model lies in an (natural) exponential family where given a functional parameter f , the random pair (Y, X) follows:

$$\begin{aligned} p_f(y, x) &= p_f(y|x)\pi(x) \\ (2.1) \quad &= \exp\{yf(x) - A(f(x)) + c(y, x)\}\pi(x), \end{aligned}$$

where $A(\cdot)$ is a known function defined upon \mathbb{R} , $c(y, x)$ is a normalizing constant, and $\pi(x)$ represents marginal density of X . For technical convenience, we assume $\underline{\pi} \leq \inf_{x \in \mathbb{I}} \pi(x) \leq \sup_{x \in \mathbb{I}} \pi(x) \leq \bar{\pi}$, for constants $\underline{\pi}, \bar{\pi} > 0$. The above framework (2.1) covers many commonly used nonparametric models; see Examples 2.1 – 2.4.

Assume that there exists a “true” parameter f_0 under which the sample is drawn from (2.1), and that f_0 belongs to an m -th order Sobolev space:

$$S^m(\mathbb{I}) = \{f \in L^2(\mathbb{I}) | f^{(j)} \text{ are abs. cont. for } j = 0, 1, \dots, m-1, \text{ and } f^{(m)} \in L^2(\mathbb{I})\}.$$

Throughout the paper, we let $m > 1/2$ such that $S^m(\mathbb{I})$ is a RKHS.

The primary model assumption in this paper is given below. Let \dot{A} , \ddot{A} , \dddot{A} be the first-, second- and third-order derivatives of A . Denote $\|f\|_\infty$ as the sup-norm of f . For any fixed $C > 0$, define $\mathcal{F}(C) = \{f \in S^m(\mathbb{I}) : \|f\|_\infty \leq C\}$.

ASSUMPTION A1. A is three-times continuously differentiable on \mathbb{R} . For any $z \in \mathbb{R}$, $\ddot{A}(z) > 0$. Moreover, for any constant $C > \|f_0\|_\infty$, there exist positive constants C_0, C_1, C_2 (possibly depending on C) such that

$$(2.2) \quad \sup_{f \in \mathcal{F}(C)} E_f \left\{ \exp(|Y - \dot{A}(f(X))|/C_0) \middle| X \right\} \leq C_1, \text{ a.s.,}$$

and for any $z \in [-2C, 2C]$,

$$(2.3) \quad 1/C_2 \leq \ddot{A}(z) \leq C_2, \text{ and } |\ddot{A}(z)| \leq C_2.$$

Assumption A1 can be easily verified in the following important examples.

EXAMPLE 2.1 (*Normal regression*). Suppose that under f , (Y, X) follows normal regression:

$$Y = f(X) + \epsilon,$$

where $\epsilon \sim N(0, 1)$. Then $A(z) = z^2/2$. For any $f \in S^m(\mathbb{I})$,

$$E_f \left\{ \exp(|Y - \dot{A}(f(X))|) \middle| X \right\} = E\{\exp(|\epsilon|)\} = \frac{2}{\sqrt{e}}(1 - \Phi(1)),$$

where $\Phi(\cdot)$ is the cumulative distribution function of ϵ . Therefore, (2.2) holds for $C_0 = 1$ and $C_1 = \frac{2}{\sqrt{e}}(1 - \Phi(1))$. It is easy to see that (2.3) holds for $C_2 = 1$.

EXAMPLE 2.2 (*Binary regression*). Suppose that under f , (Y, X) follows binary regression:

$$p_f(y|x) = \frac{\exp(f(x))}{1 + \exp(f(x))}, \text{ for } y = 0, 1.$$

Here, $A(z) = \log(1 + \exp(z))$. For any $C > \|f_0\|_\infty$ and $f \in \mathcal{F}(C)$, $|\dot{A}(f(X))| \leq (1 + \exp(-C))^{-1}$, leading to that

$$\sup_{f \in \mathcal{F}(C)} E_f \left\{ \exp(|Y - \dot{A}(f(X))|) \middle| X \right\} \leq \exp\left(\frac{2 + \exp(-C)}{1 + \exp(-C)}\right),$$

and for any $z \in [-2C, 2C]$,

$$\frac{\exp(2C)}{(1 + \exp(2C))^2} \leq \ddot{A}(z) \leq \frac{1}{4}, \text{ and } |\ddot{A}(z)| \leq \frac{1}{4},$$

which means that (2.2) holds for $C_0 = 1$ and $C_1 = \exp\left(\frac{2 + \exp(-C)}{1 + \exp(-C)}\right)$, (2.3) holds for $C_2 = \max\{\frac{1}{4}, (1 + \exp(2C))^2 \exp(-2C)\}$.

EXAMPLE 2.3 (*Binomial regression*). Suppose that under f , (Y, X) follows binomial regression:

$$p_f(y|x) = \binom{\mathfrak{a}}{y} \frac{\exp(yf(x))}{(1 + \exp(f(x)))^{\mathfrak{a}}}, \text{ for } y = 0, 1, \dots, \mathfrak{a},$$

where \mathfrak{a} is a known positive integer. In particular, $\mathfrak{a} = 1$ reduces to binary regression in Example 2.2. It is easy to see that $A(z) = \mathfrak{a} \log(1 + \exp(z))$. Similar to Example 2.2, it can be shown that (2.2) holds for $C_0 = 1$ and $C_1 = \exp\left(\frac{\mathfrak{a}+1+\exp(-C)}{1+\exp(-C)}\right)$, (2.3) holds for $C_2 = \max\{\frac{\mathfrak{a}}{4}, (1 + \exp(2C))^2 \mathfrak{a}^{-1} \exp(-2C)\}$.

EXAMPLE 2.4 (*Poisson regression*). Suppose that under f , (Y, X) follows Poisson regression:

$$p_f(y|x) = \frac{\exp(yf(x))}{y!} \exp(-\exp(f(x))), \text{ for } y = 0, 1, 2, \dots$$

Therefore, $A(z) = \exp(z)$. For any $C > \|f_0\|_\infty$ and $f \in \mathcal{F}(C)$,

$$\begin{aligned} & E_f \left\{ \exp(|Y - \dot{A}(f(X))|) \middle| X \right\} \\ & \leq \exp(\exp(C)) E_f \left\{ \exp(Y) \middle| X \right\} \\ & = \exp(\exp(C)) \times \exp(\exp(C)(e - 1)) = \exp(\exp(C)e), \end{aligned}$$

and for any $z \in [-2C, 2C]$, $\exp(-2C) \leq \ddot{A}(z) \leq \exp(2C)$ and $|\ddot{A}(z)| \leq \exp(2C)$, implying that (2.2) holds for $C_0 = 1$ and $C_1 = \exp(\exp(C)e)$, (2.3) holds for $C_2 = \exp(2C)$.

REMARK 2.1. *With stronger assumptions (e.g., stronger smoothness condition on f) and more tedious technical arguments, the results in this paper can be generalized to the following model:*

$$p_f(y|x) \sim \exp(yA_1(f(x)) - A_2(f(x)) + c(y, x)),$$

where A_1, A_2 are known functions.

Under the model Assumption A1, there exists an underlying eigen-system $(\varphi_\nu(\cdot), \rho_\nu)$ that simultaneously diagonalizes two bilinear forms V and U , where $V(g, \tilde{g}) := E\{\ddot{A}(f_0(X))g(X)\tilde{g}(X)\}$ and $U(g, \tilde{g}) := \int_0^1 g^{(m)}(x)\tilde{g}^{(m)}(x)dx$ for any $g, \tilde{g} \in S^m(\mathbb{I})$. This eigen-system forms major building blocks of constructing Gaussian processes in next section. It turns out that (φ_ν, ρ_ν) is a solution of the following ordinary differential system (whose existence and uniqueness is guaranteed by [2]):

$$\begin{aligned} & (-1)^m \varphi_\nu^{(2m)}(\cdot) = \rho_\nu \ddot{A}(f_0(\cdot)) \pi(\cdot) \varphi_\nu(\cdot), \\ (2.4) \quad & \varphi_\nu^{(j)}(0) = \varphi_\nu^{(j)}(1) = 0, \quad j = m, m+1, \dots, 2m-1, \end{aligned}$$

The following proposition summarizes some useful properties of $(\varphi_\nu(\cdot), \rho_\nu)$. Its proof can be found in [41, Proposition 2.2].

PROPOSITION 2.1. *Let Assumption A1 be satisfied. Then the sequence ρ_ν is nondecreasing satisfying $\rho_1 = \dots = \rho_m = 0$, and $\rho_\nu > 0$ for $\nu > m$. Moreover, $\rho_\nu \asymp \nu^{2m}$ and*

$$(2.5) \quad V(\varphi_\mu, \varphi_\nu) = \delta_{\mu\nu}, \quad U(\varphi_\mu, \varphi_\nu) = \rho_\mu \delta_{\mu\nu}, \quad \mu, \nu \in \mathbb{N},$$

where $\delta_{\mu\nu}$ is the Kronecker's delta. In particular, any $f \in S^m(\mathbb{I})$ admits a Fourier expansion $f = \sum_{\nu} V(f, \varphi_{\nu}) \varphi_{\nu}$ with convergence held in the $\|\cdot\|_{V,U}$ -norm¹.

3. Nonparametric Posterior Distribution. In this section, we introduce the turning prior on the infinite-dimensional parameter f . It is further demonstrated that the corresponding posterior distribution can be converted into that based on a type of penalized likelihood through a change of measure (Hájek's Lemma).

Generically, we can assume that f follows a probability measure Π_{λ} (possibly involving a hyper-parameter λ). The specification of Π_{λ} can be naturally carried out through its Radon-Nikodym (RN) derivative with respect to a base measure Π . A reasonable choice of the RN derivative is relating to the roughness penalty in smoothing spline literature ([56]):

$$(3.1) \quad \frac{d\Pi_{\lambda}}{d\Pi}(f) \propto \exp\left(-\frac{n\lambda}{2}J(f)\right),$$

where

$$J(f, g) = V(f^{\dagger}, g^{\dagger}) + U(f, g)$$

for any $f, g \in S^m(\mathbb{I})$. Here, f^{\dagger} is a projection of f onto $\mathcal{N}_m := \{g \in S^m(\mathbb{I}) : U(g, g) = 0\}$ ². Obviously, J can simultaneously control the m -order derivatives and projections onto \mathcal{N}_m through U and V . Hence, the null space of J is trivial in the sense that $J(g, g) = 0$ if and only if $g = 0$. More specifically, the prior (3.1) well controls the growth of the m -order derivative. In practice, it is indeed possible to construct Π_{λ} and Π that lead to the desired RN derivative (3.1). In this paper, we choose Π_{λ} and Π as Gaussian measures induced by a particular type of GP as discussed below. For simplicity, denote $V(g) = V(g, g)$, $U(g) = U(g, g)$ and $J(g) = J(g, g)$ from now on.

Define a GP:

$$(3.2) \quad G_{\lambda}(t) = \sum_{\nu=1}^{\infty} w_{\nu} \varphi_{\nu}(t),$$

where w_{ν} 's are independent of the observations with

$$w_{\nu} \sim \begin{cases} N(0, \sigma_{\nu}^2 / (1 + n\lambda\sigma_{\nu}^2)), & \nu = 1, 2, \dots, m, \\ N(0, 1 / (\rho_{\nu}^{1+\beta/(2m)} + n\lambda\rho_{\nu})), & \nu > m, \end{cases}$$

$\sigma_1^2, \dots, \sigma_m^2$ are fixed constants, and $\beta > 1$ is a constant. Here, the hyper-parameter β is used to measure the regularity difference between the prior and the parameter space $S^m(\mathbb{I})$; see Remark 3.1. The requirement $\beta > 1$ is necessary for G_{λ} being a valid prior on $S^m(\mathbb{I})$. In fact, if $\beta = 1$, then the path of G_{λ} does not belong to $S^m(\mathbb{I})$ almost surely (see [59, pp. 541]). However, if $\beta > 1$ and $\nu > m$, then $E\{U(G_{\lambda})\} = \sum_{\nu>m} \rho_{\nu} / (\rho_{\nu}^{1+\beta/(2m)} + n\lambda\rho_{\nu}) < \infty$, indicating that the

¹It holds that $\langle g, \tilde{g} \rangle_{V,U} = V(g, \tilde{g}) + U(g, \tilde{g})$ defines a valid inner product on $S^m(\mathbb{I})$. Let $\|\cdot\|_{V,U}$ be the corresponding norm, i.e., $\|g\|_{V,U} = \sqrt{\langle g, g \rangle_{V,U}}$.

²Specifically, \mathcal{N}_m consists of all polynomials of order up to $m-1$ ([56]).

path of G_λ almost surely belongs to $S^m(\mathbb{I})$. Let Π_λ be the probability measure induced by G_λ , i.e., $\Pi_\lambda(B) = P(G_\lambda \in B)$.

The posterior distribution of f can be written as

$$(3.3) \quad P(f|\mathbf{D}_n) \propto \exp\left(\sum_{i=1}^n [Y_i f(X_i) - A(f(X_i))]\right) d\Pi_\lambda(f),$$

where $\mathbf{D}_n \equiv \{Z_1, \dots, Z_n\}$ denotes a full sample set, and $Z_i = (Y_i, X_i)$, $i = 1, \dots, n$ are *iid* copies of $Z = (Y, X)$. By the key Lemma 3.1 below, we can link the above posterior distribution with the penalized likelihood $\ell_{n,\lambda}$ as follows:

$$(3.4) \quad P(B|\mathbf{D}_n) = \frac{\int_B \exp(n\ell_{n,\lambda}(f)) d\Pi(f)}{\int_{S^m(\mathbb{I})} \exp(n\ell_{n,\lambda}(f)) d\Pi(f)},$$

for any Π -measurable subset $B \subseteq S^m(\mathbb{I})$, where the penalized likelihood

$$\ell_{n,\lambda}(f) = \frac{1}{n} \sum_{i=1}^n [Y_i f(X_i) - A(f(X_i))] - \frac{\lambda}{2} J(f, f).$$

Here, Π is a probability measure induced by the following GP:

$$(3.5) \quad G(t) = \sum_{\nu=1}^{\infty} v_\nu \varphi_\nu(t),$$

and $\{v_\nu\}_{\nu=1}^{\infty}$ is a sequence of independent random variables (independent of \mathbf{D}_n) satisfying

$$(3.6) \quad v_\nu \sim N(0, \tau_\nu^{-2}), \text{ with } \tau_\nu^2 = \begin{cases} \sigma_\nu^{-2}, & \nu = 1, 2, \dots, m, \\ \rho_\nu^{1+\frac{\beta}{2m}}, & \nu > m. \end{cases}$$

Note that $G(\cdot)$ belongs to a family of $\mathcal{G} \equiv \{G_\lambda(\cdot) : \lambda \geq 0\}$, i.e., $\lambda = 0$. Moreover, $G(\cdot)$ can be viewed as an envelope of \mathcal{G} in the sense that their prior variances are the largest. Similarly, we can check that the path of G belongs to $S^m(\mathbb{I})$ for any $\beta > 1$ a.s.. We remark that the duality between (3.3) and (3.4) holds universally irrespective of the model assumption A1.

LEMMA 3.1. *With $f \in S^m(\mathbb{I})$, we have the following Radon-Nikodym derivative of Π_λ with respect to Π :*

$$\frac{d\Pi_\lambda}{d\Pi}(f) = \prod_{\nu=1}^m (1 + n\lambda\sigma_\nu^2)^{1/2} \prod_{\nu=m+1}^{\infty} \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \exp\left(-\frac{n\lambda}{2} J(f, f)\right).$$

REMARK 3.1. *It follows from [52] that the RKHS of G_λ is $S^{m+\frac{\beta}{2}}(\mathbb{I})$ for any $\lambda \geq 0$. The parameter space, i.e., $S^m(\mathbb{I})$, can be viewed as the completion of the above RKHS in $\|\cdot\|_{V,U}$ -norm. Similar correspondence between the parameter space and prior can be found in literature ([52, 53]).*

REMARK 3.2. *In the special Gaussian regression, the prior distributions on the frequencies of G_λ and G , i.e., the w_ν and v_ν , have a similar form with those in [56] (see Section 3.1 therein). Therefore, we can view G_λ and G as extensions of [56] to a more general framework.*

4. Posterior Contraction Rate under Sobolev Norm. In this section, we derive posterior contraction rates under a type of Sobolev norm. The Sobolev type norm is stronger than the commonly used norms such as Kullback-Leibler norm and supremum norm. Our result in this section is an intermediate step for developing BvM results, but also of independent interest.

We start from introducing a set of quantities:

$$\begin{aligned} h &= \lambda^{1/(2m)}, r_n = (nh)^{-1/2} + h^m, \tilde{r}_n = (nh)^{-1/2} + h^{m+\frac{\beta-1}{2}}, \\ D_n &= n^{-1/2} h^{-\frac{6m-1}{4m}} r_n \log n + h^{-1/2} r_n^2 \log n, \\ b_{n1} &= n^{-1/2} h^{-\frac{8m-1}{4m}} (\log n)^2 + h^{-1/2} (\log n)^{3/2}, \\ b_{n2} &= n^{-1/2} h^{-\frac{6m-1}{4m}} (\log n)^{3/2}. \end{aligned}$$

Consider a Fourier expansion $f_0(\cdot) = \sum_{\nu=1}^{\infty} f_{\nu}^0 \varphi_{\nu}(\cdot)$. We state a smoothness condition on f_0

$$\text{Condition (S): } \sum_{\nu=1}^{\infty} |f_{\nu}^0|^2 \rho_{\nu}^{1+\frac{\beta-1}{2m}} < \infty.$$

This condition is required in the quantification of the remainder terms in quadratic approximation to likelihood ratios. Heuristically, Condition (S) means that $f_0 \in S^{m+\frac{\beta-1}{2}}(\mathbb{I})$. Since the RKHS of G_{λ} is of regularity $m + \frac{\beta}{2}$ (see Remark 3.1), Condition (S) requires that the regularities of f_0 and the RKHS of G_{λ} differ by half. Such a requirement is usually needed for deriving the optimal rate of contraction; see [52].

THEOREM 4.1. *(Contraction Rate) Suppose Assumption A1 holds, and $f_0 = \sum_{\nu=1}^{\infty} f_{\nu}^0 \varphi_{\nu}$ satisfies Condition (S). Furthermore, the following Rate Condition (R) holds:*

$$\begin{aligned} r_n &= o(h^{3/2}), \quad h^{1/2} \log n = o(1), \quad nh^{2m+1} \geq 1, \quad D_n = O(\tilde{r}_n), \\ \tilde{r}_n b_{n1} &\leq 1, \quad b_{n2} \leq 1, \quad r_n^3 b_{n1} \leq \tilde{r}_n^2, \quad r_n^2 b_{n2} \leq \tilde{r}_n^2. \end{aligned}$$

Then, for any $\varepsilon_1, \varepsilon_2 \in (0, 1)$, there exist positive constants M', N' s.t. for any $n \geq N'$,

$$(4.1) \quad \mathbb{P}_{f_0}^n (P(\|f - f_0\| \geq M' \tilde{r}_n | \mathbf{D}_n) \geq \varepsilon_2) \leq \varepsilon_1,$$

where \mathbb{P}_f^n denotes the probability measure induced by \mathbf{D}_n under f .

Note that our contraction rate theorem does not require posterior consistency or posterior boundedness in contrast with [18]. The proof of Theorem 4.1 relies on Cameron-martin theorem, small ball probability, together with functional Bahadur representation ([40, 41]) and a set of empirical processes tools recently developed by [41].

REMARK 4.1. *The Rate Condition (R) characterizes the scope of the tuning parameter h (equivalently, λ). In particular, it can be shown that, when $m > 3/2$ and $1 < \beta < m + \frac{1}{2}$, Condition (R) holds for $h \asymp h^* \equiv n^{-\frac{1}{2m+\beta}}$. This leads to the contraction rate $n^{-\frac{2m+\beta-1}{2(2m+\beta)}}$. According to Condition (S), we know that f_0 belongs to the Sobolev space of higher order $S^{m+\frac{\beta-1}{2}}(\mathbb{I})$. Hence, the above contraction rate turns out to be optimal according to [20, 53] in L^2 -metric.*

REMARK 4.2. *The contraction rate derived in Theorem 4.1 is based on a Sobolev type norm, which relates to the m th order derivative of the function. Relevant results are obtained by [58] in nonparametric multivariate regression based on tensor product B-splines. In comparison, Theorem 4.1 is based on smoothing spline approach and applies to the more general exponential family.*

5. Nonparametric Bernstein-von Mises Theorem. In this section, we show that nonparametric BvM theorem holds in the exponential family considered in Section 2. To the best of our knowledge, our BvM theorem is the first one in *general* model settings based on total variation distance, extending [31] in Gaussian sequence setting. In contrast, the BvM theorems established by [9, 10] are in terms of a weaker metric: bounded Lipschitz metric. In the end, we discuss how to employ GCV to select a proper tuning prior, i.e., the value of λ .

Our BvM theorem says that the posterior measure can be well approximated by a kind of “pseudo-posterior” P_0 (in total variation distance) defined below:

$$(5.1) \quad P_0(B) = \frac{\int_B \exp(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2) d\Pi(f)}{\int_{S^m(\mathbb{I})} \exp(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2) d\Pi(f)}, \text{ for any } B \in \mathcal{B},$$

where \mathcal{B} is a collection of Π -measurable sets in $S^m(\mathbb{I})$, i.e., \mathcal{B} is a σ -algebra with respect to Π , and $\hat{f}_{n,\lambda}$ is a smoothing spline estimate defined as

$$(5.2) \quad \hat{f}_{n,\lambda} = \arg \max_{f \in S^m(\mathbb{I})} \ell_{n,\lambda}(f).$$

In spite of its complicated appearance, P_0 turns out to be a Gaussian measure on $S^m(\mathbb{I})$ induced by a Gaussian process W , which has a closed form (5.4). This Gaussianity facilitates the applications of our BvM theorem in the next section.

THEOREM 5.1. *(Nonparametric BvM theorem) Suppose Assumption A1 holds, and $f_0 = \sum_{\nu=1}^{\infty} f_{\nu}^0 \varphi_{\nu}$ satisfies Condition (S). Furthermore, let Condition (R) be satisfied and the b_{n1}, b_{n2} therein further satisfy, as $n \rightarrow \infty$, $n\tilde{r}_n^2(b_{n1} + b_{n2}) = o(1)$. Then we have, as $n \rightarrow \infty$,*

$$(5.3) \quad \sup_{B \in \mathcal{B}} |P(B|\mathbf{D}_n) - P_0(B)| = o_{P_{f_0}^n}(1).$$

Theorem 5.1 is the main result of our paper. So, we sketch its proof here. According to Theorem 4.1, the posterior mass is mostly concentrated on an $M\tilde{r}_n$ -ball of f_0 for a large M , denoted as $\mathbb{B}_{M\tilde{r}_n}(f_0)$. Hence, for any $B \in \mathcal{B}$, we decompose $P(B|\mathbf{D}_n) = P(B \cap \mathbb{B}_{M\tilde{r}_n}(f_0)|\mathbf{D}_n) + P(B \cap \mathbb{B}_{M\tilde{r}_n}^c(f_0)|\mathbf{D}_n)$. Following Theorem 4.1, the second term is uniformly negligible for all $B \in \mathcal{B}$. Applying Taylor expansion to the penalized likelihood in (3.4) (in terms of Fréchet derivatives), we can show that the first term is asymptotically close to $P_0(B)$ uniformly for $B \in \mathcal{B}$ based on a series of empirical processes techniques.

Conditions in Theorem 5.1 are not restrictive at all. Indeed, by direct calculations, we can verify that $h \asymp h^* \equiv n^{-\frac{1}{2m+\beta}}$ satisfies Condition (R) and $n\tilde{r}_n^2(b_{n1} + b_{n2}) = o(1)$ when $m > 1 + \frac{\sqrt{3}}{2} \approx 1.866$

and $1 < \beta < m + \frac{1}{2}$. Therefore, (5.3) holds under $h \asymp h^*$. Interestingly, the optimal posterior contraction rate is simultaneously obtained in this case; see Remark 4.1.

We next show that P_0 is equivalent to a Gaussian measure Π_W , induced by a Gaussian process W defined in (5.4). As will be seen later, nonparametric Bayesian inference procedures in Section 6 can be easily conducted based on W . Suppose that $\hat{f}_{n,\lambda}$ satisfies the following Fourier expansion:

$$\hat{f}_{n,\lambda}(\cdot) = \sum_{\nu=1}^{\infty} \hat{f}_{\nu} \varphi_{\nu}(\cdot)$$

Let

$$(5.4) \quad W(\cdot) = \sum_{\nu=1}^{\infty} (a_{n,\nu} \hat{f}_{\nu} + b_{n,\nu} \tau_{\nu} v_{\nu}) \varphi_{\nu}(\cdot),$$

where $a_{n,\nu} = n(1 + \lambda \gamma_{\nu})(\tau_{\nu}^2 + n(1 + \lambda \gamma_{\nu}))^{-1}$, $b_{n,\nu} = (\tau_{\nu}^2 + n(1 + \lambda \gamma_{\nu}))^{-1/2}$, v_{ν} and τ_{ν}^2 satisfy (3.6), and the sequence γ_{ν} is defined as

$$(5.5) \quad \gamma_{\nu} = \begin{cases} 1, & \nu = 1, 2, \dots, m, \\ \rho_{\nu}, & \nu > m. \end{cases}$$

Define $\tilde{f}_{n,\lambda}(\cdot) = \sum_{\nu=1}^{\infty} a_{n,\nu} \hat{f}_{\nu} \varphi_{\nu}(\cdot)$. Remark that $\tilde{f}_{n,\lambda} \neq \hat{f}_{n,\lambda}$. Clearly, we can re-express W as

$$W = \tilde{f}_{n,\lambda} + W_n,$$

where $W_n(\cdot) := \sum_{\nu=1}^{\infty} b_{n,\nu} \tau_{\nu} v_{\nu} \varphi_{\nu}(\cdot)$ is a mean-zero GP. Hence, W is centering around $\tilde{f}_{n,\lambda}$.

Let Π_W be the probability measure induced by W . Theorem 5.2 below essentially says that $P_0(B|\mathbf{D}_n) = P(W \in B|\mathbf{D}_n) := \Pi_W(B)$ for any $B \in \mathcal{B}$. Together with Theorem 5.1, we note that the posterior distribution $P(\cdot|\mathbf{D}_n)$ and $\Pi_W(\cdot)$ are asymptotically close under the total variation distance. Hence, the mean function $\tilde{f}_{n,\lambda}$ is approximately the posterior mode of $P(\cdot|\mathbf{D}_n)$. Therefore, $\tilde{f}_{n,\lambda}$ can be used as the center of credible regions to be constructed.

THEOREM 5.2. *With $f \in S^m(\mathbb{I})$, the Radon-Nikodym derivative of Π_W with respect to Π is*

$$\frac{d\Pi_W}{d\Pi}(f) = \frac{\exp(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2)}{\int_{S^m(\mathbb{I})} \exp(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2) d\Pi(f)}.$$

Hence, we have

$$\frac{dP_0}{d\Pi}(f) = \frac{d\Pi_W}{d\Pi}(f).$$

In the end of this section, we point out a practical implication of the duality between nonparametric Bayesian models and smoothing spline models. Specifically, we can employ the well developed GCV method to select a proper tuning prior, i.e., the value of λ (equivalently, h), in practice. Let h_{GCV} be the GCV-selected tuning parameter based on the penalized likelihood function, and h_{GCV} is known to achieve the approximate optimal rate $n^{-1/(2m+1)}$ (see [55]). As

illustrated in next section, h should be chosen in the order of $n^{-\frac{1}{2m+\beta}}$ for credible regions/intervals to possess frequentist validity. Hence, we set

$$h = h_{GCV}^{(2m+1)/(2m+\beta)}$$

in the tuning prior. Numerical evidences in Section 7 strongly support this new method.

6. Nonparametric Bernstein-von Mises Phenomenon. Nonparametric BvM theorem can be applied to construct credible regions and credible intervals for linear functionals, as illustrated in this section. Our results are classified into two types: finite sample construction and asymptotic construction. It should be noted that a variety of nonparametric models including Examples 2.1–2.4 are covered in this section. The hyper-parameter λ will be shown to control the frequentist performance of these Bayesian inference procedures.

Throughout this section, we suppose that f_0 satisfies Condition (S), and let $h \asymp h^* = n^{-\frac{1}{2m+\beta}}$. Recall that such a selection of h yields an optimal contraction rate $n^{-\frac{2m+\beta-1}{2(2m+\beta)}}$ (Remark 4.1).

6.1. Credible Region in Strong Topology. As the first application of Theorem 5.1, we consider the construction of credible region for f in terms of L^2 -norm, and also study its frequentist property. Relevant results include [31, 27, 9, 28, 47, 10, 35, 48] in Gaussian white noise, [44, 45] in Gaussian regression with fixed design, and [58] in Gaussian regression with sieved priors. In contrast, our results are established in the more general nonparametric exponential family.

For any $f \in S^m(\mathbb{I})$, define $\|f\|_2 = V(f)^{1/2}$, a type of L^2 -norm. For any $\alpha \in (0, 1)$, let $r_n(\alpha) > 0$ satisfy $P(f \in S^m(\mathbb{I}) : \|f - \tilde{f}_{n,\lambda}\|_2 \leq r_n(\alpha) | \mathbf{D}_n) = 1 - \alpha$. Hence, a credible region with credibility level $(1 - \alpha)$ is

$$(6.1) \quad R_n(\alpha) = \left\{ f \in S^m(\mathbb{I}) : \|f - \tilde{f}_{n,\lambda}\|_2 \leq r_n(\alpha) \right\}.$$

We next examine the frequentist property of $R_n(\alpha)$.

THEOREM 6.1. *Suppose that Assumption A1 holds, f_0 satisfies Condition (S), $m > 1 + \frac{\sqrt{3}}{2}$, $1 < \beta < m + 1/2$, and $h \asymp h^*$. Then for any $\alpha \in (0, 1)$, $\lim_{n \rightarrow \infty} P_{f_0}^n(f_0 \in R_n(\alpha)) = 1$.*

6.2. Credible Region in Weak Topology. The frequentist coverage of the credible region (6.1) asymptotically approaches one regardless of the credibility level (see Theorem 6.1). This motivates us to construct a modified credible region using a *weaker* topology such that the truth can be covered with probability approaching exactly the credibility level. Our proof relies on the general BvM theorem in Section 5 and a strong approximation result ([49]). In [9, 10], similar results were obtained under Gaussian white noise model and density estimation. Our result can be viewed as a nontrivial extension to nonparametric exponential family.

For any $f \in S^m(\mathbb{I})$ with $f(\cdot) = \sum_{\nu=1}^{\infty} f_{\nu} \varphi_{\nu}(\cdot)$, define $\|f\|_{\omega}^2 = \sum_{\nu=1}^{\infty} \omega_{\nu} f_{\nu}^2$, where ω_{ν} is a given positive sequence satisfying $\omega_{\nu} = \nu^{-1}(\log 2\nu)^{-\tau}$ for a constant $\tau > 1$. Since $\omega_{\nu} \leq 1$ for all $\nu \geq 1$,

it is easy to see that $\|f\|_\omega \leq \|f\|_2$. Therefore, $\|\cdot\|_\omega$ is weaker than $\|\cdot\|_2$. We will show that under this weaker norm, any $(1 - \alpha)$ credible region can recover *exactly* $(1 - \alpha)$ frequentist coverage.

For any $\alpha \in (0, 1)$, let $r_{\omega,n}(\alpha) > 0$ satisfy $P(f \in S^m(\mathbb{I}) : \|f - \tilde{f}_{n,\lambda}\|_\omega \leq r_{\omega,n}(\alpha) | \mathbf{D}_n) = 1 - \alpha$. We construct a credible region with credibility level $(1 - \alpha)$:

$$(6.2) \quad R_n^\omega(\alpha) = \left\{ f \in S^m(\mathbb{I}) : \|f - \tilde{f}_{n,\lambda}\|_\omega \leq r_{\omega,n}(\alpha) \right\}.$$

Theorem 6.2 below proves that the credible region $R_n^\omega(\alpha)$ asymptotically matches with the frequentist coverage.

THEOREM 6.2. *Suppose that Assumption A1 holds, f_0 satisfies Condition (S), $m > 1 + \frac{\sqrt{3}}{2}$, $1 < \beta < \min\{m + \frac{1}{2}, \frac{(2m-1)^2}{2m}\}$, and $h \asymp h^*$. Then for any $\alpha \in (0, 1)$, $\lim_{n \rightarrow \infty} P_{f_0}^n(f_0 \in R_n^\omega(\alpha)) = 1 - \alpha$.*

The L^2 -diameter of $R_n^\omega(\alpha)$ is infinity (see [42, Section S.5]). But we can impose a restriction such that the L^2 -diameter becomes finite, by using a strategy of [9]. Specifically, define

$$R_n^{*\omega}(\alpha) = R_n^\omega(\alpha) \cap \{f \in S^m(\mathbb{I}) : J(f) \leq M\},$$

for a constant $M > 0$. It can be shown that the L^2 -diameter of $R_n^{*\omega}(\alpha)$ is $O_{P_{f_0}^n}(n^{-\frac{2m+\beta-1}{2(2m+\beta)}} \sqrt{\log n})$ (see [42, Section S.5]). Recall the leading factor $n^{-\frac{2m+\beta-1}{2(2m+\beta)}}$ is the optimal contraction rate under Sobolev norm (see Remark 4.1). So the L^2 -diameter of $R_n^{*\omega}(\alpha)$ is the optimal contraction rate multiplying a root logarithmic factor.

6.3. Linear Functionals on the Regression Function. In this section, we apply Theorem 5.1 to construct credible intervals for a general class of linear functionals in nonparametric exponential family. Frequentist coverage of the proposed credible interval is also investigated. In particular, we consider two important special cases: (i) evaluation functional: $F_z(f) = f(z)$, where $z \in \mathbb{I}$ is a fixed number; (ii) integral functional: $F_\omega(f) = \int_0^1 f(z)\omega(z)dz$, where $\omega(\cdot)$ is a known deterministic integrable function such as an indicator function. We find that the former leads to an interval contracting at slower than root- n rate, while the latter leads to root- n rate.

The existing literature mostly focus on functionals where efficient estimation with \sqrt{n} -rate is available ([39, 9, 10, 11]). The more general inefficient estimation with slower than root- n rate (e.g., evaluation functional) is only treated recently by [46] in Gaussian white noise model. As will be seen, our theory treat efficient and inefficient functionals in a unified framework.

Let $F : S^m(\mathbb{I}) \mapsto \mathbb{R}$ be a linear Π -measurable functional, i.e., $F(af + bg) = aF(f) + bF(g)$ for any $a, b \in \mathbb{R}$ and $f, g \in S^m(\mathbb{I})$. We say that F satisfies Condition (F) if there exist constants $\kappa > 0$ and $r \in [0, 1]$ such that for any $f \in S^m(\mathbb{I})$,

$$(6.3) \quad |F(f)| \leq \kappa h^{-r/2} \|f\|.$$

Lemma 6.3 below (given in [41]) implies that both F_z and F_ω satisfy (6.3).

LEMMA 6.3. *There exists a universal constant $c > 0$ s.t. for any $f \in S^m(\mathbb{I})$, $\|f\|_\infty \leq ch^{-1/2}\|f\|$.*

Let $r_{F,n}(\alpha) > 0$ satisfy $P(f \in S^m(\mathbb{I}) : |F(f) - F(\tilde{f}_{n,\lambda})| \leq r_{F,n}(\alpha) | \mathbf{D}_n) = 1 - \alpha$. Define $(1 - \alpha)$ credible interval for $F(f)$ as

$$(6.4) \quad CI_n^F(\alpha) : F(\tilde{f}_{n,\lambda}) \pm r_{F,n}(\alpha).$$

Theorem 6.4 below shows that CI_n^F covers the true value $F(f_0)$ with probability asymptotically at least $1 - \alpha$ for a general class of functionals. For $k \geq 1$, define

$$\theta_{k,n}^2 = \sum_{\nu=1}^{\infty} \frac{F(\varphi_\nu)^2}{(\tau_\nu^2 + n(1 + \lambda\gamma_\nu))^k}.$$

THEOREM 6.4. *Suppose that Assumption A1 holds, $f_0 = \sum_{\nu=1}^{\infty} f_\nu^0 \varphi_\nu$ satisfies Condition (S') : $\sum_{\nu=1}^{\infty} |f_\nu^0|^2 \nu^{2m+\beta} < \infty$, $m > 1 + \frac{\sqrt{3}}{2}$, $1 < \beta < \min\{m + \frac{1}{2}, \frac{(2m-1)^2}{2m}\}$, and $h \asymp h^*$. Meanwhile,*

$$(6.5) \quad n^k \theta_{k,n}^2 \asymp h^{-r} \text{ for } k = 1, 2.$$

Then for any $\alpha \in (0, 1)$,

$$(6.6) \quad \liminf_{n \rightarrow \infty} \mathbf{P}_{f_0}^n(F(f_0) \in CI_n^F(\alpha)) \geq 1 - \alpha,$$

given that Condition (F) holds. Moreover, if $0 < \sum_{\nu=1}^{\infty} F(\varphi_\nu)^2 < \infty$, then $\lim_{n \rightarrow \infty} \mathbf{P}_{f_0}^n(F(f_0) \in CI_n^F(\alpha)) = 1 - \alpha$.

Remark that Condition (S') is slightly stronger than Condition (S).

Condition (6.5) is not restrictive and can be verified in concrete settings, e.g., when model is Gaussian (Example 2.1) and F is evaluation functional or integral functional. We summarize this result in the following Proposition 6.5. The proof relies on a nice closed form of φ_ν and a careful analysis of the trigonometric functions.

PROPOSITION 6.5. *Suppose $m = 2$, $X \sim \text{Unif}[0, 1]$, and $Y|f, X \sim N(f(X), 1)$.*

- (i) *If $F = F_z$ for any $z \in (0, 1)$, then (6.5) holds for $r = 1$;*
- (ii) *If $F = F_\omega$ for any $\omega \in L^2(\mathbb{I}) \setminus \{0\}$, then $0 < \sum_{\nu=1}^{\infty} F(\varphi_\nu)^2 < \infty$ and (6.5) holds for $r = 0$.*

REMARK 6.1. *The radii $r_n(\alpha)$, $r_{\omega,n}(\alpha)$ and $r_{F,n}(\alpha)$ are determined by posterior samples of f . This might be time-consuming in practice. In fact, the proof of Theorems 6.1, 6.2 and 6.4 reveal that the radii satisfy the following large-sample (data-free) limits:*

$$(6.7) \quad \begin{aligned} r_n(\alpha) &\approx \sqrt{\frac{\zeta_{1,n} + \sqrt{2\zeta_{2,n}z_\alpha}}{n}}, \text{ where } \zeta_{k,n} = \sum_{\nu=1}^{\infty} \frac{1}{(1 + \lambda\gamma_\nu + n^{-1}\tau_\nu^2)^k}, \\ r_{\omega,n}(\alpha) &\approx \sqrt{\frac{c_\alpha}{n}}, \\ r_{F,n}(\alpha) &\approx \theta_{1,n} z_{\alpha/2}, \end{aligned}$$

where $c_\alpha > 0$ satisfies $\mathbf{P}(\sum_{\nu=1}^{\infty} \omega_\nu \eta_\nu^2 \leq c_\alpha) = 1 - \alpha$ with η_ν being independent standard normal random variables, and $z_\alpha = \Phi^{-1}(1 - \alpha)$ with Φ being the standard Gaussian c.d.f. Replacing the radii by the above limits (6.7), one can establish the asymptotic proxies of (6.1), (6.2) and (6.4), which can reduce computational burden. The frequentist coverage of these asymptotic regions/intervals can be shown to be the same as the original ones.

REMARK 6.2. By carefully examining the proof of Theorem 6.4, we find that when $F = F_z$, the inequality (6.6) is actually strict, and the length of $CI_n^F(\alpha)$ satisfies $r_{F,n}(\alpha) \asymp \theta_{1,n} \asymp n^{-\frac{2m+\beta-1}{2(2m+\beta)}}$ (Remark 6.1). However, when $F = F_\omega$, $CI_n^F(\alpha)$ covers the truth with probability approaching $1 - \alpha$, and the length of $CI_n^F(\alpha)$ satisfies $r_{F,n}(\alpha) \asymp \theta_{1,n} \asymp n^{-1/2}$. Therefore, we observe a subtle difference between the two types of functionals. Simulation results in Section 7 further demonstrate their empirical distinctions.

7. Simulations. In this section, we empirically investigate the frequentist coverage probabilities of the credible region (6.1) and modified credible region (6.2), and credible intervals for evaluation functional and integral functional.

We generated data from the following model

$$(7.1) \quad Y_i = f_0(X_i) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

where X_i are *iid* uniform over $[0, 1]$, and ϵ_i are *iid* standard normal random variables independent of X_i . The true regression function f_0 was chosen as $f_0(x) = 3\beta_{30,17}(x) + 2\beta_{3,11}(x)$, where $\beta_{a,b}$ is the probability density function for $Beta(a, b)$. Figure 1 displays the true function f_0 , from which it can be seen that f_0 has both peaks and troughs.

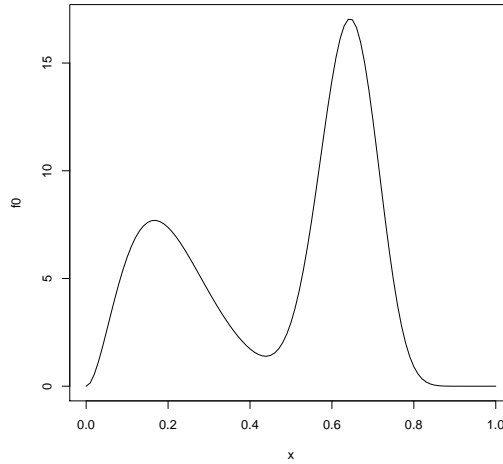


FIG 1. Plot of the true function f_0 used in model (7.1).

To examine the coverage property of the credible regions, we chose n ranging from 20 to 2000. For each n , 1,000 independent trials were conducted. From each trial, a credible region (CR) based on (6.1) and a modified credible region (MCR) based on (6.2) were constructed. Proportions of the CR and MCR covering f_0 were calculated, and were displayed against the sample sizes. Results are summarized in Figure 2. It can be seen that for different $1 - \alpha$, i.e., the credibility levels, the coverage proportions (CP) of CR are greater than $1 - \alpha$ when n is large enough. They even tend to one for large sample sizes. However, the CP of the MCR tends to exactly $1 - \alpha$ when n increases. Thus, the numerical results confirm our theory developed in Sections 6.1 and 6.2.

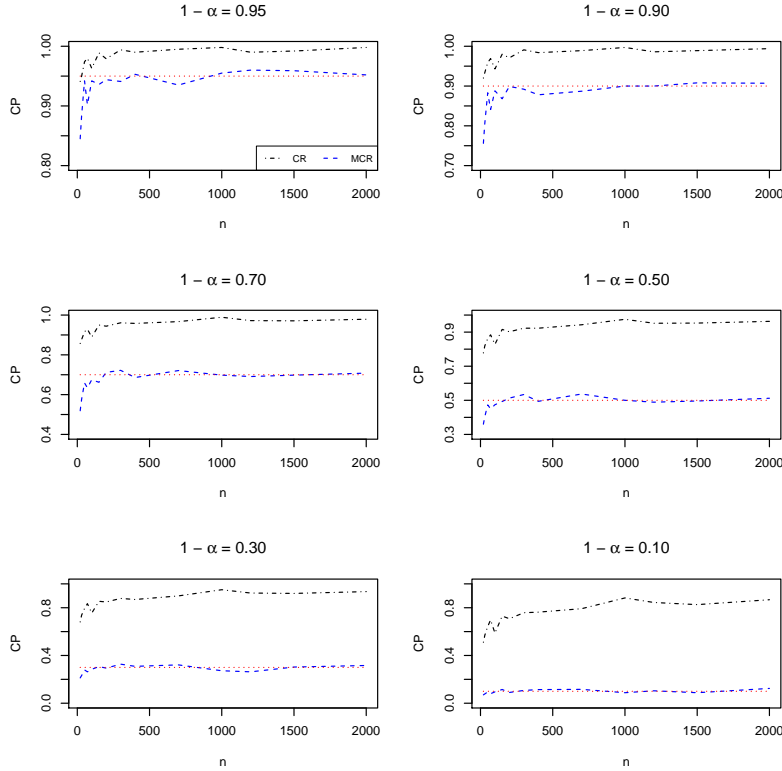


FIG 2. Coverage proportion (CP) of CR and MCR, constructed by (6.1) and (6.2) respectively, based on various sample sizes and credibility levels. The dotted red line indicates the position of the $1 - \alpha$ credibility level.

To examine the coverage property of credible intervals, we chose $n = 2^5, 2^7, 2^8, 2^9$ to demonstrate the trend of coverage along with increasing sample sizes. For evaluation functional, we considered $F = F_z$ for 15 evenly-spaced z points in $[0, 1]$. For each z , a credible interval based on (6.4) was constructed. We then calculated the coverage probability of this interval based on 1,000 independent experiments, that is, the empirical proportion of the intervals (among the 1,000 intervals) that cover the true value $f_0(z)$. Figure 3 summarizes the results for different credibility levels α , where coverage probabilities are plotted against the corresponding points z . It can be seen that the coverage probability of the pointwise intervals is a bit larger than $1 - \alpha$ for all α

and n being considered. This is consistent with Proposition 6.5 (i), except for the points near the right peak of f_0 . Indeed, at those points near the right peak, under-coverage has been observed. This is a common phenomenon in the frequentist literature: the peak and trouts may affect the coverage property of the pointwise interval; see [36, 41].

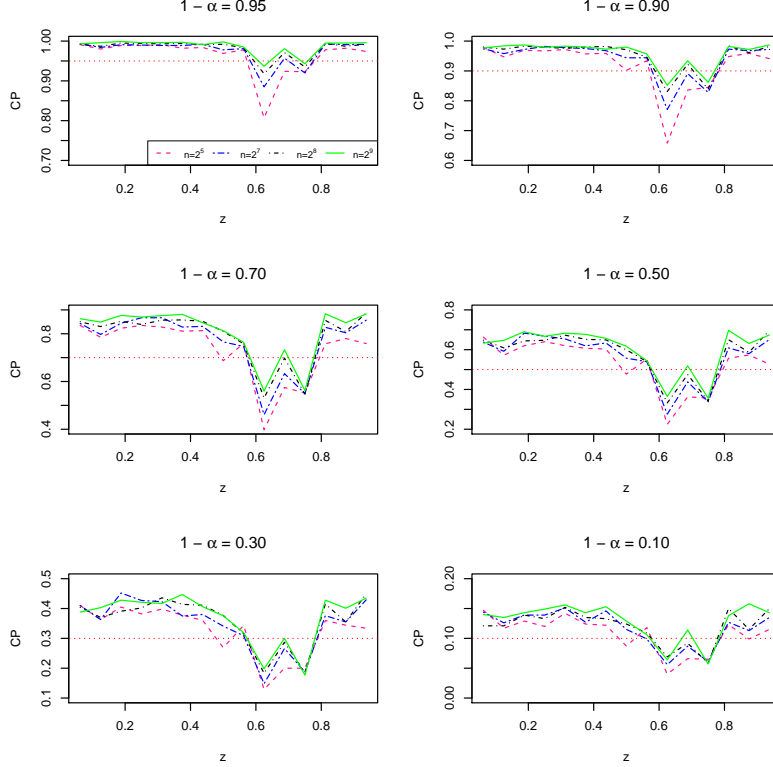


FIG 3. Coverage proportion (CP) of the credible interval for $F_z(f_0)$ versus z . The dotted red line indicates the position of the $1 - \alpha$ credibility level.

For integral functional, we considered $F = F_{\omega_{z_0}}$ for $\omega_{z_0}(z) = I(0 \leq z \leq z_0)$ with 15 evenly-spaced z_0 points in $[0, 1]$. We evaluated the coverage probability at each z_0 based on 1,000 experiments. Figure 4 summarizes the results for different credibility levels α , where coverage probabilities are plotted against the corresponding points z_0 . It can be seen that, as n increases, the coverage probability of the integral intervals tends to $1 - \alpha$ for all α . This phenomenon is consistent with our theory, i.e., Proposition 6.5 (ii).

Acknowledgements. We thank Prof. Jayanta Ghosh for careful reading and comments, and also thank PhD student Meimei Liu at Purdue for help with the simulation study.

8. APPENDIX. In this section, we will prove the main results: Theorems 5.2, 6.1, 6.2 and 6.4. Due to limited space, proofs of Theorems 4.1 and 5.1 are deferred to supplementary document [42].

The following result is a special case of [42, Proposition S.1] which will be useful for future

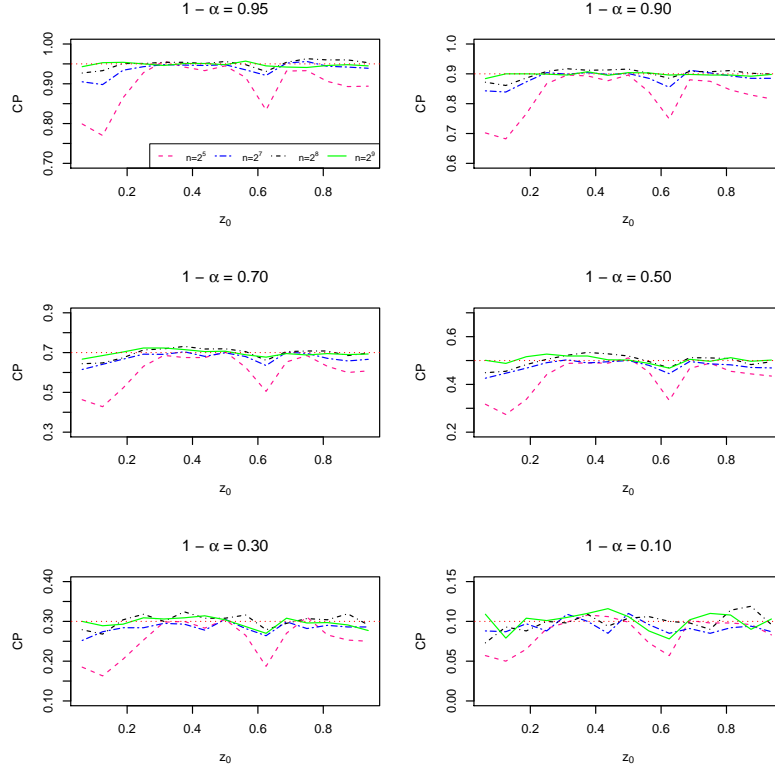


FIG 4. Coverage proportion (CP) of the credible interval for $F_{\omega_{z_0}}(f_0)$ versus z_0 . The dotted red line indicates the position of the $1 - \alpha$ credibility level.

proofs.

Proposition A.1. For any $f \in S^m(\mathbb{I})$ and $z \in \mathbb{I}$, we have $\|f\|^2 = \sum_{\nu} |V(f, \varphi_{\nu})|^2 (1 + \lambda \rho_{\nu})$, $K_z(\cdot) = \sum_{\nu} \frac{\varphi_{\nu}(z)}{1 + \lambda \rho_{\nu}} \varphi_{\nu}(\cdot)$, and $W_{\lambda} \varphi_{\nu}(\cdot) = \frac{\lambda \rho_{\nu}}{1 + \lambda \rho_{\nu}} \varphi_{\nu}(\cdot)$ under Assumption 2.1.

PROOF OF THEOREM 5.2. For any $f \in S^m(\mathbb{I})$, by Proposition 2.1, f admits a unique series representation $f = \sum_{\nu=1}^{\infty} f_{\nu} \varphi_{\nu}$, where $f_{\nu} = V(f, \varphi_{\nu})$ satisfies $\sum_{\nu} f_{\nu}^2 \rho_{\nu} < \infty$. Therefore, $T : f \mapsto \{f_{\nu} : \nu \geq 1\}$ defines a one-to-one map from $S^m(\mathbb{I})$ to $\mathcal{R}_m \equiv \{\{f_{\nu}\}_{\nu=1}^{\infty} \in \mathbb{R}^{\infty} : \sum_{\nu=1}^{\infty} f_{\nu}^2 \rho_{\nu} < \infty\}$.

Let Π'_W and Π' be the probability measures induced by $\{a_{n,\nu} \hat{f}_{\nu} + b_{n,\nu} \tau_{\nu} v_{\nu} : \nu \geq 1\}$ and $\{v_{\nu} : \nu \geq 1\}$. Then $d\Pi'_W/d\Pi'$ equals $\lim_{N \rightarrow \infty} p_{1,N}(f_1, \dots, f_N)/p_{2,N}(f_1, \dots, f_N)$ (see [43, Section III]), where $p_{1,N}$ and $p_{2,N}$ are the probability densities under $f_{\nu} \sim a_{n,\nu} \hat{f}_{\nu} + b_{n,\nu} \tau_{\nu} v_{\nu}$ and $f_{\nu} \sim v_{\nu}$, $\nu = 1, \dots, N$, respectively. A direct evaluation leads to that

$$\frac{d\Pi'_W}{d\Pi'}(\{f_{\nu} : \nu \geq 1\}) = C_{n,\lambda} \exp\left(-\frac{n}{2} \sum_{\nu=1}^{\infty} (f_{\nu} - \hat{f}_{\nu})^2 (1 + \lambda \gamma_{\nu})\right),$$

where

$$C_{n,\lambda} = \prod_{\nu=1}^{\infty} (b_{n,\nu} \tau_{\nu})^{-1} \exp\left(\frac{n}{2} \sum_{\nu=1}^{\infty} \frac{\tau_{\nu}^2 (1 + \lambda \gamma_{\nu})}{\tau_{\nu}^2 + n(1 + \lambda \gamma_{\nu})} \hat{f}_{\nu}^2\right).$$

Since $\sum_{\nu} \widehat{f}_{\nu}^2 \gamma_{\nu} < \infty$ and $\beta > 1$, it is not hard to see that $C_{n,\lambda}$ is an almost surely finite constant.

For any $B \subseteq \mathcal{R}_m$, $\Pi_W(T^{-1}B) = \Pi'_W(B)$, and $\Pi(T^{-1}B) = \Pi'(B)$. To see this, note that

$$\Pi_W(T^{-1}B) = \mathbb{P}(W \in T^{-1}B) = \mathbb{P}(\{a_{n,\nu}\widehat{f}_{\nu} + b_{n,\nu}\tau_{\nu}v_{\nu} : \nu \geq 1\} \in B) = \Pi'_W(B), \text{ and}$$

$$\Pi(T^{-1}B) = \mathbb{P}(G \in T^{-1}B) = \mathbb{P}(\{v_{\nu} : \nu \geq 1\} \in B) = \Pi'(B).$$

By change of variable, for any Π -measurable $S \subseteq S^m(\mathbb{I})$,

$$\begin{aligned} \Pi_W(S) &= \Pi'_W(TS) \\ &= \int_{TS} d\Pi'_W(\{f_{\nu} : \nu \geq 1\}) \\ &= C_{n,\lambda} \int_{TS} \exp\left(-\frac{n}{2} \sum_{\nu=1}^n (f_{\nu} - \widehat{f}_{\nu})^2 (1 + \lambda\gamma_{\nu})\right) d\Pi'(\{f_{\nu} : \nu \geq 1\}) \\ &= C_{n,\lambda} \int_S \exp\left(-\frac{n}{2} \|f - \widehat{f}_{n,\lambda}\|^2\right) d\Pi(f). \end{aligned}$$

In particular, let $S = S^m(\mathbb{I})$ in the above equations, we get that

$$C_{n,\lambda} = \left(\int_{S^m(\mathbb{I})} \exp\left(-\frac{n}{2} \|f - \widehat{f}_{n,\lambda}\|^2\right) d\Pi(f) \right)^{-1}.$$

This proves the desired result. \square

Before proving Theorem 6.1, we give some preliminary results.

Lemma A.1. As $n \rightarrow \infty$, $\frac{n\|W_n\|_2^2 - \zeta_{1,n}}{\sqrt{2\zeta_{2,n}}} \xrightarrow{d} N(0, 1)$.

PROOF OF LEMMA A.1. Let $\eta_{\nu} = \tau_{\nu}v_{\nu}$. Then η_{ν} is a sequence of *iid* standard normals. Note that

$$\|W_n\|_2^2 = \sum_{\nu=1}^{\infty} \frac{\eta_{\nu}^2}{\tau_{\nu}^2 + n(1 + \lambda\gamma_{\nu})}.$$

Let $U_n = (n\|W_n\|_2^2 - \zeta_{1,n})/\sqrt{2\zeta_{2,n}}$, then we have

$$U_n = \frac{1}{\sqrt{2\zeta_{2,n}}} \sum_{\nu=1}^{\infty} \frac{n(\eta_{\nu}^2 - 1)}{\tau_{\nu}^2 + n(1 + \lambda\gamma_{\nu})}.$$

By straightforward calculations and Taylor's expansion of $\log(1 - x)$, it can be shown that the logarithm of the moment generating function of U_n equals

$$(A.1) \quad \log E\{\exp(tU_n)\} = t^2/2 + O\left(t^3 \zeta_{2,n}^{-3/2} \zeta_{3,n}\right).$$

It can be examined that $\zeta_{2,n} \asymp n^{1/(2m+\beta)}$ and $\zeta_{3,n} \asymp n^{1/(2m+\beta)}$, so the remainder term in (A.1) is $O(n^{-1/(2(2m+\beta))}) = o(1)$. So $\lim_{n \rightarrow \infty} E\{\exp(tU_n)\} = \exp(t^2/2)$. Proof is completed. \square

Define

$$(A.2) \quad R(x, y) = \sum_{\nu=1}^{\infty} \frac{\varphi_{\nu}(x)\varphi_{\nu}(y)}{1 + \lambda\gamma_{\nu} + n^{-1}\tau_{\nu}^2}, \quad x, y \in \mathbb{I}.$$

Lemma A.2. $\sup_{x,y \in \mathbb{I}} |R(x, y)| \lesssim h^{-1}$ and $\sup_{x,y \in \mathbb{I}} \left| \frac{\partial}{\partial x} R(x, y) \right| \lesssim h^{-2}$.

PROOF OF LEMMA A.2. For any $g \in S^m(\mathbb{I})$ and $x \in \mathbb{I}$, it follows from [16, Lemmas (2.10) and (2.17)] that there exist constants c', c'', c''' s.t.

$$\begin{aligned} \left| \left\langle g, \frac{\partial}{\partial x} K_x \right\rangle \right| &= \left| \frac{\partial}{\partial x} \langle g, K_x \rangle \right| \\ &= |g'(x)| \leq c' h^{-1/2} \sqrt{\|g'\|_{L^2}^2 + h^2 \|g''\|_{L^2}^2} \\ &= c' h^{-3/2} \sqrt{h^2 \|g'\|_{L^2}^2 + h^4 \|g''\|_{L^2}^2} \\ &\leq c' h^{-3/2} \sqrt{(\|g\|_{L^2}^2 + h^2 \|g'\|_{L^2}^2) + (\|g\|_{L^2}^2 + h^4 \|g''\|_{L^2}^2)} \\ &\leq c' c'' h^{-3/2} \sqrt{\|g\|_{L^2}^2 + h^{2m} \|g^{(m)}\|_{L^2}^2} \\ &\leq c''' h^{-3/2} \|g\|. \end{aligned}$$

This implies that $\|\frac{\partial}{\partial x} K_x\| \leq c''' h^{-3/2}$. For convenience, let $R_y(\cdot) = R(\cdot, y)$. It is easy to see that

$$\begin{aligned} \|R_y\|^2 &= \sum_{\nu=1}^{\infty} \frac{\varphi_{\nu}(y)^2}{(1 + \lambda\gamma_{\nu} + n^{-1}\tau_{\nu}^2)^2} (1 + \lambda\gamma_{\nu}) \\ &\leq \sum_{\nu=1}^{\infty} \frac{\varphi_{\nu}(y)^2}{1 + \lambda\gamma_{\nu}} = K(y, y) \leq c_K^2 h^{-1}. \end{aligned}$$

This implies that $|R(x, y)| = |\langle R_y, K_x \rangle| \leq \|R_y\| \cdot \|K_x\| \leq c_K^2 h^{-1}$. This also leads to that, for any $x, y \in \mathbb{I}$,

$$\left| \frac{\partial}{\partial x} R(x, y) \right| = \left| \left\langle R_y, \frac{\partial}{\partial x} K_x \right\rangle \right| \leq \|R_y\| \cdot \left\| \frac{\partial}{\partial x} K_x \right\| \leq c''' c_K h^{-2}.$$

The desired result follows by the fact that both c_k and c''' are universal constants free of x, y . \square

PROOF OF THEOREM 6.1. By direct examinations, we can show that the Rate Conditions **(R)**, $n\tilde{r}_n^2(\tilde{r}_n b_{n1} + b_{n2}) = o(1)$, $nh^{1/2}D_n^2 = o(1)$ are all satisfied.

It is sufficient to investigate the $\mathbf{P}_{f_0}^n$ -probability of the event $\|\tilde{f}_{n,\lambda} - f_0\|_2 \leq r_n(\alpha)$. To achieve this goal, we first prove the following fact:

$$(A.3) \quad |z_n(\alpha) - z_{\alpha}| = o_{\mathbf{P}_{f_0}^n}^n(1),$$

where $z_{\alpha} = \Phi^{-1}(1 - \alpha)$ and Φ is the c.d.f. of $N(0, 1)$, and $z_n(\alpha) = (nr_n(\alpha)^2 - \zeta_{1,n})/\sqrt{2\zeta_{2,n}}$. The proof of the theorem follows by (A.3) and a careful analysis of $f_0 - \tilde{f}_{N,\lambda}$.

We first show (A.3). It follows by Theorem 5.1 that,

$$|P(R_n(\alpha)|\mathbf{D}_n) - P_0(R_n(\alpha))| \leq \sup_{B \in \mathcal{B}} |P(B|\mathbf{D}_n) - P_0(B)| = o_{\mathbf{P}_{f_0}^n}^n(1).$$

Together with $P(R_n(\alpha)|\mathbf{D}_n) = 1 - \alpha$, we have

$$|P_0(R_n(\alpha)) - (1 - \alpha)| = o_{\mathbf{P}_{f_0}^n}(1).$$

Since $W = \tilde{f}_{n,\lambda} + W_n$,

$$\begin{aligned} P_0(R_n(\alpha)) &= \mathbf{P}(W \in R_n(\alpha)|\mathbf{D}_n) \\ &= \mathbf{P}(\|W_n\|_2 \leq r_n(\alpha)|\mathbf{D}_n) = \mathbf{P}(U_n \leq z_n(\alpha)|\mathbf{D}_n), \end{aligned}$$

and $\mathbf{P}(U_n \leq z_\alpha) \rightarrow 1 - \alpha$, where U_n is defined in the proof of Lemma A.1, we get that

$$(A.4) \quad |\mathbf{P}(U_n \leq z_n(\alpha)|\mathbf{D}_n) - \mathbf{P}(U_n \leq z_\alpha)| = o_{\mathbf{P}_{f_0}^n}(1),$$

where $U_n = (n\|W_n\|_2^2 - \zeta_{1,n})/\sqrt{2\zeta_{2,n}}$. Let Φ_n be the c.d.f. of U_n . Since U_n is independent of the data, we have from (A.4) that

$$(A.5) \quad |\Phi_n(z_n(\alpha)) - \Phi_n(z_\alpha)| = o_{\mathbf{P}_{f_0}^n}(1).$$

Now for any $\varepsilon > 0$, if $|z_n(\alpha) - z_\alpha| \geq \varepsilon$, then either $|\Phi_n(z_n(\alpha)) - \Phi_n(z_\alpha)| \geq \Phi_n(z_\alpha + \varepsilon) - \Phi_n(z_\alpha)$ or $|\Phi_n(z_n(\alpha)) - \Phi_n(z_\alpha)| \geq \Phi_n(z_\alpha) - \Phi_n(z_\alpha - \varepsilon)$. Since Φ_n pointwise converges to Φ , both $\Phi_n(z_\alpha + \varepsilon) - \Phi_n(z_\alpha)$ and $\Phi_n(z_\alpha) - \Phi_n(z_\alpha - \varepsilon)$ are asymptotically lower bounded by some positive numbers (possibly depending on ε). This implies by (A.5) that (A.3) holds.

Next we prove the theorem. Define $Rem_n = \hat{f}_{n,\lambda} - f_0 - S_{n,\lambda}(f_0)$. It follows by *Functional Bahadur Representation* ([41, Theorem 3.4]), or equivalently, equation (S.26) in supplement [42] that $\|Rem_n\| = O_{\mathbf{P}_{f_0}^n}(D_n)$ with $D_n = a_n + b_n$. By direct examination, we have

$$\begin{aligned} &\tilde{f}_{n,\lambda} - f_0 \\ &= \sum_{\nu=1}^{\infty} \left(a_{n,\nu} V(\hat{f}_{n,\lambda}, \varphi_\nu) - f_\nu^0 \right) \varphi_\nu \\ &= \sum_{\nu=1}^{\infty} \left(a_{n,\nu} V(Rem_n + f_0 + S_{n,\lambda}(f_0), \varphi_\nu) - f_\nu^0 \right) \varphi_\nu \\ &= \sum_{\nu=1}^{\infty} a_{n,\nu} V(Rem_n, \varphi_\nu) \varphi_\nu + \sum_{\nu=1}^{\infty} (a_{n,\nu} - 1) f_\nu^0 \varphi_\nu \\ (A.6) \quad &+ \sum_{\nu=1}^{\infty} a_{n,\nu} V\left(\frac{1}{n} \sum_{i=1}^n \epsilon_i K_{X_i}, \varphi_\nu\right) \varphi_\nu - \sum_{\nu=1}^{\infty} a_{n,\nu} V(\mathcal{P}_\lambda f_0, \varphi_\nu) \varphi_\nu, \end{aligned}$$

where $\epsilon_i = Y_i - \dot{A}(f_0(X_i))$. Denote the four terms in the above equation by T_1, T_2, T_3, T_4 .

Since $a_{n,\nu} \leq 1$, it is easy to see that

$$\begin{aligned} \|T_1\|_2^2 &= \sum_{\nu=1}^{\infty} a_{n,\nu}^2 |V(Rem_n, \varphi_\nu)|^2 \\ (A.7) \quad &\leq \sum_{\nu=1}^{\infty} |V(Rem_n, \varphi_\nu)|^2 = \|Rem_n\|_2^2 \leq \|Rem_n\|^2 = O_{\mathbf{P}_{f_0}^n}(D_n^2). \end{aligned}$$

Using $h \asymp n^{-1/(2m+\beta)}$ and a direct algebra we get that

$$\begin{aligned} \|T_2\|_2^2 &= \sum_{\nu=1}^{\infty} (a_{n,\nu} - 1)^2 |f_\nu^0|^2 \\ &\asymp \sum_{\nu=1}^{\infty} \left(\frac{\nu^{2m+\beta}}{\nu^{2m+\beta} + n(1 + \lambda\nu^{2m})} \right)^2 |f_\nu^0|^2 \\ &= o(n^{-\frac{2m+\beta-1}{2m+\beta}}) = o(n^{-1}h^{-1}). \end{aligned}$$

Meanwhile, it follows by Proposition A.1 that

$$\begin{aligned} \|T_4\|_2^2 &= \sum_{\nu=1}^{\infty} a_{n,\nu}^2 |f_\nu^0|^2 \left(\frac{\lambda\gamma_\nu}{1 + \lambda\gamma_\nu} \right)^2 \\ &\leq \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \left(\frac{\lambda\gamma_\nu}{1 + \lambda\gamma_\nu} \right)^2 \\ &\lesssim \sum_{\nu=1}^{\infty} |f_\nu^0|^2 (h\nu)^{2m+\beta-1} \frac{(h\nu)^{2m-\beta+1}}{(1 + (h\nu)^{2m})^2} \\ &= o(n^{-\frac{2m+\beta-1}{2m+\beta}}) = o(n^{-1}h^{-1}). \end{aligned}$$

It is easy to see that $R(x, x') = \sum_{\nu=1}^{\infty} a_{n,\nu} \frac{\varphi_\nu(x)\varphi_\nu(x')}{1 + \lambda\gamma_\nu}$ for any $x, x' \in \mathbb{I}$. Also define $R_x(\cdot) = R(x, \cdot)$. It is easy to see that $R_x \in S^m(\mathbb{I})$ for any $x \in \mathbb{I}$. Then it can be shown that $T_3 = \frac{1}{n} \sum_{i=1}^n \epsilon_i R_{X_i}$, leading to

$$\|T_3\|_2^2 = V(T_3, T_3) = \frac{1}{n^2} \sum_{i=1}^n \epsilon_i^2 V(R_{X_i}, R_{X_i}) + \frac{2}{n^2} \sum_{1 \leq i < k \leq n} \epsilon_i \epsilon_k V(R_{X_i}, R_{X_k}).$$

Define $W(n) = 2 \sum_{i < k} \epsilon_i \epsilon_k V(R_{X_i}, R_{X_k})$. Let $W_{ik} = 2\epsilon_i \epsilon_k V(R_{X_i}, R_{X_k})$ for $1 \leq i < k \leq n$, then $W(n) = \sum_{1 \leq i < k \leq n} W_{ik}$. Note that $W(n)$ is clean in the sense of [15]. Let $\sigma^2(n) = E_{f_0}\{W(n)^2\}$ and G_I, G_{II}, G_{IV} be defined as

$$\begin{aligned} G_I &= \sum_{1 \leq i < j \leq n} E_{f_0}\{W_{ij}^4\}, \\ G_{II} &= \sum_{1 \leq i < j < k \leq n} (E_{f_0}\{W_{ij}^2 W_{ik}^2\} + E_{f_0}\{W_{ji}^2 W_{jk}^2\} + E_{f_0}\{W_{ki}^2 W_{kj}^2\}), \text{ and} \\ G_{IV} &= \sum_{1 \leq i < j < k < l \leq n} (E_{f_0}\{W_{ij} W_{ik} W_{lj} W_{lk}\} + E_{f_0}\{W_{ij} W_{il} W_{kj} W_{kl}\} \\ &\quad + E_{f_0}\{W_{ik} W_{il} W_{jk} W_{jl}\}). \end{aligned}$$

Since φ_ν are uniformly bounded, we get that

$$\|R_x\|_2^2 = \sum_{\nu=1}^{\infty} \frac{|\varphi_\nu(x)|^2}{(1 + n^{-1}\tau_\nu^2 + \lambda\gamma_\nu)^2} \lesssim h^{-1},$$

where “ \lesssim ” is free of x . This implies that $G_I = O(n^2 h^{-4})$ and $G_{II} = O(n^3 h^{-4})$. It can also be shown that for pairwise distinct i, k, t, l ,

$$\begin{aligned} & E_{f_0}\{W_{ik}W_{il}W_{tk}W_{tl}\} \\ &= 2^4 E_{f_0}\{\epsilon_i^2 \epsilon_k^2 \epsilon_t^2 \epsilon_l^2 V(R_{X_i}, R_{X_k})V(R_{X_i}, R_{X_l})V(R_{X_t}, R_{X_k})V(R_{X_t}, R_{X_l})\} \\ &= 2^4 \sum_{\nu=1}^{\infty} \frac{a_{n,\nu}^8}{(1 + \lambda \gamma_\nu)^8} = O(h^{-1}), \end{aligned}$$

which implies that $G_{IV} = O(n^4 h^{-1})$. In the mean time, a straight algebra leads to that

$$\begin{aligned} \sigma^2(n) &= 4 \binom{n}{2} \sum_{\nu=1}^{\infty} \frac{a_{n,\nu}^4}{(1 + \lambda \gamma_\nu)^4} \\ &= 4 \binom{n}{2} \sum_{\nu=1}^{\infty} \left(\frac{n}{\tau_\nu^2 + n(1 + \lambda \gamma_\nu)} \right)^4 = 2n(n-1)\zeta_{4,n} \asymp n^2 h^{-1}. \end{aligned}$$

Since $nh^2 \asymp n^{1-2/(2m+\beta)} \rightarrow \infty$, we get that G_I, G_{II} and G_{IV} are all of order $o(\sigma^4(n))$. Then it follows by [15] that as $n \rightarrow \infty$,

$$\frac{W(n)}{n\sqrt{2\zeta_{4,n}}} \xrightarrow{d} N(0, 1).$$

Since $\zeta_{4,n} \asymp h^{-1}$, the above equation leads to that $W(n)/n = O_{\mathbf{P}_{f_0}^n}(h^{-1/2})$. It follows by direct examination that

$$\text{Var}_{f_0}\left\{\sum_{i=1}^n \epsilon_i^2 V(R_{X_i}, R_{X_i})\right\} \leq n E_{f_0}\{\epsilon_i^4 \|R_{X_i}\|_2^4\} = O(nh^{-2}),$$

leading to that

$$\begin{aligned} \sum_{i=1}^n \epsilon_i^2 V(R_{X_i}, R_{X_i}) &= E_{f_0}\left\{\sum_{i=1}^n \epsilon_i^2 V(R_{X_i}, R_{X_i})\right\} + O_{\mathbf{P}_{f_0}^n}(n^{1/2}h^{-1}) \\ &= n\zeta_{2,n} + O_{\mathbf{P}_{f_0}^n}(n^{1/2}h^{-1}). \end{aligned}$$

Therefore, it follows by condition $nhD_n^2 = o(1)$ and the above analysis on T_1, T_2, T_3, T_4 that

$$\begin{aligned} nh\|\tilde{f}_{n,\lambda} - f_0\|_2^2 &= nh\|T_3\|_2^2 + O_{\mathbf{P}_{f_0}^n}(nhD_n^2) + o_{\mathbf{P}_{f_0}^n}(1) \\ (A.8) \quad &= h\zeta_{2,n} + o_{\mathbf{P}_{f_0}^n}(1). \end{aligned}$$

In the end, note from (A.3) and $\zeta_{k,n} \asymp n^{1/(2m+\beta)}$ (see the proof of Lemma A.1) that

$$nr_n(\alpha)^2 = \zeta_{1,n} + \sqrt{2\zeta_{2,n}}z_\alpha + o_{\mathbf{P}_{f_0}^n}(\sqrt{\zeta_{2,n}}).$$

Therefore,

$$nhr_n(\alpha)^2 = h\zeta_{1,n}(1 + o_{\mathbf{P}_{f_0}^n}(1)).$$

Since $\liminf_{n \rightarrow \infty}(h\zeta_{1,n} - h\zeta_{2,n}) > 0$, we get that, with $\mathbf{P}_{f_0}^n$ -probability approaching one, $\|\tilde{f}_{n,\lambda} - f_0\|_2 \leq r_n(\alpha)$, i.e., $f_0 \in R_n(\alpha)$. Proof is completed. \square

Before proving Theorem 6.2, let us present two preliminary lemmas.

Lemma A.3. *As $n \rightarrow \infty$,*

$$n\|W_n\|_\omega^2 \xrightarrow{d} \sum_{\nu=1}^{\infty} \omega_\nu \eta_\nu^2,$$

where η_ν are independent standard normal random variables.

PROOF OF LEMMA A.3. The proof follows by moment generating function approach and direct calculations, as in the proof of Lemma A.1. \square

PROOF OF THEOREM 6.2. By direct examinations, one can show that Rate Conditions (R), $n\tilde{r}_n^2(\tilde{r}_n b_{n1} + b_{n2}) = o(1)$, and $nD_n^2 = o(1)$ are all satisfied.

We first have the following fact:

$$(A.9) \quad |\sqrt{n}r_{\omega,n}(\alpha) - \sqrt{c_\alpha}| = o_{\mathbf{P}_{f_0}^n}(1),$$

where $c_\alpha > 0$ satisfies $\mathbf{P}(\sum_{\nu=1}^{\infty} \omega_\nu \eta_\nu^2 \leq c_\alpha) = 1 - \alpha$ with η_ν being independent standard normal random variables. It follows from (A.9) that

$$(A.10) \quad nr_{\omega,n}(\alpha)^2 = c_\alpha + o_{\mathbf{P}_{f_0}^n}(1).$$

The proof of (A.9) is similar to the proof of (A.3) and is omitted.

Let T_1, T_2, T_3, T_4 be items defined in (A.6). It follows from the proof of Theorem 6.1 that

$$\|T_1\|_\omega^2 \leq \|T_1\|_2^2 = O_{\mathbf{P}_{f_0}^n}(D_n^2),$$

so $n\|T_1\|_\omega^2 = O_{\mathbf{P}_{f_0}^n}(nD_n^2) = o_{\mathbf{P}_{f_0}^n}(1)$ due to the condition $nD_n^2 = o(1)$.

It follows by condition $h \asymp n^{-1/(2m+\beta)}$, dominated convergence theorem and direct examinations that

$$\begin{aligned} \|T_2\|_\omega^2 &= \sum_{\nu=1}^{\infty} \omega_\nu (a_{n,\nu} - 1)^2 |f_\nu^0|^2 \\ &\asymp n^{-2} \sum_{\nu=1}^{\infty} \omega_\nu \frac{\nu^{2m+\beta+1}}{(1 + (h\nu)^{2m} + (h\nu)^{2m+\beta})^2} \times \nu^{2m+\beta-1} |f_\nu^0|^2 \\ &\lesssim n^{-1} \sum_{\nu=1}^{\infty} \frac{(h\nu)^{2m+\beta+1}}{(1 + (h\nu)^{2m} + (h\nu)^{2m+\beta})^2} \times \nu^{2m+\beta-1} |f_\nu^0|^2 = o(n^{-1}), \end{aligned}$$

and

$$\begin{aligned} \|T_4\|_\omega^2 &= \sum_{\nu=1}^{\infty} \omega_\nu a_{n,\nu}^2 \left(\frac{\lambda \gamma_\nu}{1 + \lambda \gamma_\nu} \right)^2 |f_\nu^0|^2 \\ &\lesssim \sum_{\nu=1}^{\infty} \omega_\nu \frac{(h\nu)^{2m-\beta+1}}{(1 + (h\nu)^{2m} + (h\nu)^{2m+\beta})^2} \times |f_\nu^0|^2 (h\nu)^{2m+\beta-1} \\ &\lesssim h^{2m+\beta} \sum_{\nu=1}^{\infty} \frac{(h\nu)^{2m-\beta}}{(1 + (h\nu)^{2m} + (h\nu)^{2m+\beta})^2} \times |f_\nu^0|^2 \nu^{2m+\beta-1} = o(n^{-1}). \end{aligned}$$

Next we handle T_3 . By proof of Theorem 6.1, we have $T_3 = n^{-1} \sum_{i=1}^n \epsilon_i R_{X_i}$, where $\epsilon_i = Y_i - \dot{A}(f_0(X_i))$, which implies $n\|T_3\|_\omega^2 = n^{-1} \|\sum_{i=1}^n \epsilon_i R_{X_i}\|_\omega^2$.

Since $E_{f_0}\{\exp(|\epsilon|/C_0)\} \leq C_1$, we can choose a constant $L > C_0$ such that $P_{f_0}^n(\mathcal{E}_n) \rightarrow 1$, where $\mathcal{E}_n = \{\max_{1 \leq i \leq n} |\epsilon_i| \leq b_n \equiv L \log n\}$. We can even choose the above L to be properly large so that the following rate condition holds:

$$(A.11) \quad h^{-1}n^{1/2} \exp(-b_n/(2C_0)) = o(1), \quad h^{-2} \exp(-b_n/(2C_0)) = o(1).$$

Define

$$H_n(\cdot) = n^{-1/2} \sum_{i=1}^n \epsilon_i R_{X_i}(\cdot), \quad H_n^b(\cdot) = n^{-1/2} \sum_{i=1}^n \epsilon_i I(|\epsilon_i| \leq b_n) R_{X_i}(\cdot).$$

Write

$$H_n = H_n - H_n^b - E_{f_0}\{H_n - H_n^b\} + H_n^b - E_{f_0}\{H_n^b\}.$$

Clearly, on \mathcal{E}_n , $H_n = H_n^b$, and hence,

$$\begin{aligned} & |H_n(z) - H_n^b(z) - E_{f_0}\{H_n(z) - H_n^b(z)\}| \\ &= |E_{f_0}\{H_n(z) - H_n^b(z)\}| \\ &= n^{1/2} |E_{f_0}\{\epsilon I(|\epsilon| > b_n) R_X(z)\}| \\ &\lesssim n^{1/2} h^{-1} E_{f_0}\{\epsilon^2\}^{1/2} \mathbf{P}_{f_0}^n(|\epsilon| > b_n)^{1/2} \\ &\lesssim n^{1/2} h^{-1} \exp(-b_n/(2C_0)) = o(1), \end{aligned}$$

where the last $o(1)$ -term follows by (A.11) and is free of the argument z . Thus,

$$(A.12) \quad \sup_{z \in \mathbb{I}} |H_n(z) - H_n^b(z) - E_{f_0}\{H_n(z) - H_n^b(z)\}| = o_{\mathbf{P}_{f_0}^n}(1).$$

Define $\mathcal{R}_n = H_n^b - E_{f_0}\{H_n^b\}$ and $Z_n(e, x) = n^{1/2}(P_n(e, x) - P(e, x))$, where $P_n(e, x)$ is the empirical distribution of (ϵ, X) and $P(e, x)$ is the population distribution of (ϵ, X) under $\mathbf{P}_{f_0}^n$ -probability. It follows by Theorem 1 of [49] that

$$(A.13) \quad \sup_{e \in \mathbb{R}, x \in \mathbb{I}} |Z_n(e, x) - W(t(e, x))| = O_{\mathbf{P}_{f_0}^n}(n^{-1/2}(\log n)^2),$$

where $W(\cdot, \cdot)$ is Brownian bridge indexed on \mathbb{I}^2 , $t(e, x) = (F_1(x), F_2(e|x))$, F_1 is the marginal distribution of X and F_2 is the conditional distribution of ϵ given X both under $\mathbf{P}_{f_0}^n$ -probability. It can be seen that

$$\mathcal{R}_n(z) = \int_0^1 \int_{-b_n}^{b_n} e R_x(z) dZ_n(e, x).$$

Define

$$\mathcal{R}_n^0(z) = \int_0^1 \int_{-b_n}^{b_n} e R_x(z) dW(t(e, x)).$$

Write

$$dU_n(x) = \int_{-b_n}^{b_n} e dZ_n(e, x), \quad dU_n^0(x) = \int_{-b_n}^{b_n} e dW(t(e, x)).$$

It follows from integration by parts where all quadratic variation terms are zero that

$$U_n(x) = \int_{-b_n}^{b_n} e d_e Z_n(e, x) = Z_n(e, x) e|_{e=-b_n}^{b_n} - \int_{-b_n}^{b_n} Z_n(e, x) de,$$

$$U_n^0(x) = \int_{-b_n}^{b_n} e d_e W(t(e, x)) = W(t(e, x)) e|_{e=-b_n}^{b_n} - \int_{-b_n}^{b_n} W(t(e, x)) de,$$

and hence, it follows by (A.13) that

$$\sup_{x \in \mathbb{I}} |U_n(x) - U_n^0(x)| = O_{\mathbb{P}_{f_0}^n}(b_n n^{-1/2} (\log n)^2).$$

It follows from integration by parts again and $\sup_{x, y \in \mathbb{I}} |\frac{\partial}{\partial x} R(x, y)| = O(h^{-2})$ (Lemma A.2) that

$$\mathcal{R}_n(z) = \int_0^1 R_x(z) dU_n(x) = U_n(x) R(x, z)|_{x=0}^1 - \int_0^1 U_n(x) \frac{\partial}{\partial x} R(x, z) dx,$$

$$\mathcal{R}_n^0(z) = \int_0^1 R_x(z) dU_n^0(x) = U_n^0(x) R(x, z)|_{x=0}^1 - \int_0^1 U_n^0(x) \frac{\partial}{\partial x} R(x, z) dx,$$

and hence,

$$(A.14) \quad \sup_{z \in \mathbb{I}} |\mathcal{R}_n(z) - \mathcal{R}_n^0(z)| = O_{\mathbb{P}_{f_0}^n}(h^{-2} b_n n^{-1/2} (\log n)^2) = o_{\mathbb{P}_{f_0}^n}(1),$$

where the last equality follows by $2m + \beta > 4$ and hence

$$h^{-2} n^{-1/2} b_n (\log n)^2 = O(n^{-1/2+2/(2m+\beta)} (\log n)^3) = o(1).$$

Next we handle the term \mathcal{R}_n^0 . Write $W(s, t) = B(s, t) - stB(1, 1)$, where $B(s, t)$ is standard Brownian motion indexed on \mathbb{I}^2 . Define

$$\bar{\mathcal{R}}_n^0(z) = \int_0^1 \int_{-b_n}^{b_n} e R_x(z) dB(t(e, x)).$$

Let $F(e, x) = F_1(x)F_2(e, x)$ be the joint distribution of (ϵ, X) . It is easy to see that

$$\begin{aligned} |\bar{\mathcal{R}}_n^0(z) - \mathcal{R}_n^0(z)| &= |B(1, 1)| \cdot \left| \int_0^1 \int_{-b_n}^{b_n} e R_x(z) dF(e, x) \right| \\ &= |B(1, 1)| \cdot |E_{f_0}\{\epsilon I(|\epsilon| \leq b_n) R_X(z)\}| \\ &= |B(1, 1)| \cdot |E_{f_0}\{\epsilon I(|\epsilon| > b_n) R_X(z)\}| \\ &= O_{\mathbb{P}_{f_0}^n}(h^{-1} \exp(-b_n/(2C_0))) = o_{\mathbb{P}_{f_0}^n}(1), \end{aligned}$$

where the last equality follows by (A.11). Therefore, we have shown that

$$(A.15) \quad \sup_{z \in \mathbb{I}} |\bar{\mathcal{R}}_n^0(z) - \mathcal{R}_n^0(z)| = o_{\mathbb{P}_{f_0}^n}(1).$$

By (A.12), (A.14) and (A.15) that

$$(A.16) \quad n \|T_3\|_\omega^2 = \|H_n\|_\omega^2 = \|\bar{\mathcal{R}}_n^0\|_\omega^2 + o_{\mathbb{P}_{f_0}^n}(1).$$

Define

$$\tilde{\mathcal{R}}(z) = \int_0^1 \int_{-\infty}^{\infty} e R_x(z) dB(t(e, x)).$$

Let $\Delta(z) = \tilde{\mathcal{R}}(z) - \bar{\mathcal{R}}_n^0(z)$. Then

$$\Delta(z) = \int_0^1 \int_{|e| > b_n} e R_x(z) dB(t(e, x)).$$

For each z , $\Delta(z)$ is a zero-mean Gaussian random variable with variance

$$\begin{aligned} E_{f_0} \{\Delta(z)^2\} &= \int_0^1 \int_{|e| > b_n} e^2 R_x(z)^2 dF(e, x) \\ &\lesssim h^{-2} E_{f_0} \{\epsilon^2 I(|e| > b_n)\} = O(h^{-2} \exp(-b_n/(2C_0))) = o(1), \end{aligned}$$

where the last $o(1)$ -term follows from (A.11) and is free of the argument z . Therefore,

$$E_{f_0} \{\|\tilde{\mathcal{R}} - \bar{\mathcal{R}}_n^0\|_{\omega}^2\} \lesssim E_{f_0} \{\|\Delta\|_{L^2}^2\} = \int_0^1 E_{f_0} \{\Delta(z)^2\} dz = o(1),$$

implying that $\|\tilde{\mathcal{R}} - \bar{\mathcal{R}}_n^0\|_{\omega} = o_{\mathbf{P}_{f_0}^n}(1)$. Therefore, it follows by (A.16) that

$$(A.17) \quad n\|T_3\|_{\omega}^2 = \|\tilde{\mathcal{R}}\|_{\omega}^2 + o_{\mathbf{P}_{f_0}^n}(1).$$

It follows from the definition of $R(\cdot, \cdot)$ that

$$\|\tilde{\mathcal{R}}\|_{\omega}^2 = \sum_{\nu=1}^{\infty} \frac{\omega_{\nu} \tilde{\eta}_{\nu}^2}{(1 + \lambda \gamma_{\nu} + n^{-1} \tau_{\nu}^2)^2} \xrightarrow{d} \sum_{\nu=1}^{\infty} \omega_{\nu} \tilde{\eta}_{\nu}^2,$$

where $\tilde{\eta}_{\nu} = \int_0^1 \int_{-\infty}^{\infty} e \varphi_{\nu}(x) dB(t(e, x))$. It is easy to see that for any ν, μ ,

$$E_{f_0} \{\tilde{\eta}_{\nu} \tilde{\eta}_{\mu}\} = E_{f_0} \{\epsilon^2 \varphi_{\nu}(X) \varphi_{\mu}(X)\} = E_{f_0} \{B(X) \varphi_{\nu}(X) \varphi_{\mu}(X)\} = V(\varphi_{\nu}, \varphi_{\mu}) = \delta_{\nu\mu},$$

that is, $\tilde{\eta}_{\nu}$ are *iid* standard normal random variables. Combined with the above analysis of terms T_1, T_2, T_3, T_4 , we have shown that as $n \rightarrow \infty$,

$$n\|f_0 - \tilde{f}_{n,\lambda}\|_{\omega}^2 \xrightarrow{d} \sum_{\nu=1}^{\infty} \omega_{\nu} \tilde{\eta}_{\nu}^2.$$

This implies that as $n \rightarrow \infty$,

$$\mathbf{P}_{f_0}^n(f_0 \in R_n^{\omega}(\alpha)) = \mathbf{P}_{f_0}^n(n\|f_0 - \tilde{f}_{n,\lambda}\|_{\omega}^2 \leq c_{\alpha}) \rightarrow 1 - \alpha.$$

The proof is completed. \square

PROOF OF THEOREM 6.4. Recall in the proof of Theorem 6.2 we show that Rate Conditions **(R)**, $n\tilde{r}_n^2(\tilde{r}_n b_{n1} + b_{n2}) = o(1)$ and $nD_n^2 = o(1)$ are all satisfied.

It is easy to see that

$$(A.18) \quad F(W_n) \stackrel{d}{=} N(0, \theta_{1,n}^2).$$

Define $R_n^F(\alpha) = \{f \in S^m(\mathbb{I}) : |F(f) - F(\tilde{f}_{n,\lambda})| \leq r_{F,n}(\alpha)\}$. It follows by Theorem 5.1 that $|1 - \alpha - P_0(R_n^F(\alpha))| = o_{\mathbf{P}_{f_0}^n}(1)$. It is easy to see that

$$P_0(R_n^F(\alpha)) = \mathbf{P}(|F(W_n)| \leq r_{F,n}(\alpha) | \mathbf{D}_n) = 2\Phi(r_{F,n}(\alpha)/\theta_{1,n}) - 1,$$

which leads to

$$(A.19) \quad |r_{F,n}(\alpha)/\theta_{1,n} - z_{\alpha/2}| = o_{\mathbf{P}_{f_0}^n}(1).$$

Consider the decomposition (A.6) with T_1, T_2, T_3, T_4 being defined therein. It follows by (A.7) and rate condition $nD_n^2 = o(1)$ that $n\|T_1\|^2 = O_{P_{f_0}}(nD_n^2) = o_{\mathbf{P}_{f_0}^n}(1)$. Meanwhile, it follows by Condition (S'), $n^{-1} \asymp h^{2m+\beta}$ and $\lambda = h^{2m}$ and direct examinations that

$$\begin{aligned} n\|T_2\|^2 &= n \sum_{\nu=1}^{\infty} (a_{n,\nu} - 1)^2 |f_\nu^0|^2 (1 + \lambda\gamma_\nu) \\ &\asymp n \sum_{\nu=1}^{\infty} \left(\frac{\nu^{2m+\beta}}{\nu^{2m+\beta} + n(1 + \lambda\nu^{2m})} \right)^2 |f_\nu^0|^2 (1 + \lambda\nu^{2m}) \\ &\asymp \sum_{\nu=1}^{\infty} \frac{(h\nu)^{2m+\beta} + (h\nu)^{4m+\beta}}{(1 + (h\nu)^{2m} + (h\nu)^{2m+\beta})^2} \times |f_\nu^0|^2 \nu^{2m+\beta} = o(1), \end{aligned}$$

and

$$\begin{aligned} n\|T_4\|^2 &= n \sum_{\nu=1}^{\infty} a_{n,\nu}^2 \left(\frac{\lambda\gamma_\nu}{1 + \lambda\gamma_\nu} \right)^2 |f_\nu^0|^2 (1 + \lambda\gamma_\nu) \\ &\asymp \sum_{\nu=1}^{\infty} \frac{(h\nu)^{2m-\beta}}{1 + (h\nu)^{2m}} \times |f_\nu^0|^2 \nu^{2m+\beta} = o(1). \end{aligned}$$

Therefore, $\|\tilde{f}_{n,\lambda} - f_0 - T_3\| = \|T_1 + T_2 + T_4\| = o_{\mathbf{P}_{f_0}^n}(n^{-1/2})$. It follows from (6.3) that $|F(\tilde{f}_{n,\lambda} - f_0) - F(T_3)| = o_{\mathbf{P}_{f_0}^n}(h^{-r/2}n^{-1/2})$.

Recall $F(T_3) = \frac{1}{n} \sum_{i=1}^n \epsilon_i F(R_{X_i})$, where the kernel R_x is defined in (A.2). We will derive the asymptotic distribution for $F(T_3)$. Let $s_n^2 = \text{Var}_{f_0}(\sum_{i=1}^n \epsilon_i F(R_{X_i}))$. It is easy to show that

$$s_n^2 = n^3 \sum_{\nu=1}^{\infty} \frac{F(\varphi_\nu)^2}{(\tau_\nu^2 + n(1 + \lambda\gamma_\nu))^2} = n^3 \theta_{2,n}^2 \asymp nh^{-r}.$$

By (6.3) and $\|R_x\| \leq c_K h^{-1/2}$ (see proof of Lemma A.2), we get

$$|F(R_x)| \leq \kappa h^{-r/2} \|R_x\| \leq \kappa c_K h^{-(1+r)/2}.$$

Meanwhile,

$$(A.20) \quad E_{f_0}\{\epsilon^2 F(R_X)^2\} = n^2 \sum_{\nu=1}^{\infty} \frac{F(\varphi_\nu)^2}{(\tau_\nu^2 + n(1 + \lambda\gamma_\nu))^2} = n^2 \theta_{2,n}^2 \asymp h^{-r}.$$

By Assumption A1 there exists a constant M_4 s.t. $E_{f_0}\{\epsilon^4|X\} \leq M_4$ a.s. Then for any $\delta > 0$,

$$\begin{aligned} & \frac{1}{s_n^2} \sum_{i=1}^n E_{f_0}\{\epsilon_i^2 F(R_{X_i})^2 I(|\epsilon_i F(R_{X_i})| \geq \delta s_n)\} \\ & \leq \frac{n}{s_n^2} (\delta s_n)^{-2} E_{f_0}\{\epsilon^4 F(R_X)^4\} \\ & \lesssim \frac{n}{s_n^2} (\delta s_n)^{-2} h^{-(1+r)} E_{f_0}\{\epsilon^2 F(R_X)^2\} \lesssim \delta^{-2} n^{-1} h^{-1} = o(1), \end{aligned}$$

where the last $o(1)$ -term follows by $h \asymp h^*$ and $2m + \beta > 1$. By Lindeberg's central limit theorem, as $n \rightarrow \infty$,

$$(A.21) \quad \frac{F(T_3)}{\sqrt{n}\theta_{2,n}} = \frac{1}{s_n} \sum_{i=1}^n \epsilon_i F(R_{X_i}) \xrightarrow{d} N(0, 1).$$

By condition $n^2 \theta_{2,n}^2 \asymp h^{-r}$, we have

$$\left| \frac{F(\tilde{f}_{n,\lambda} - f_0 - T_3)}{\sqrt{n}\theta_{2,n}} \right| = o_{\mathbf{P}_{f_0}^n} \left(\frac{h^{-r/2} n^{-1/2}}{\sqrt{n}\theta_{2,n}} \right) = o_{\mathbf{P}_{f_0}^n}(1).$$

It follows by (A.19) that

$$\frac{r_{F,n}(\alpha)}{\sqrt{n}\theta_{2,n}} = \frac{\theta_{1,n}}{\sqrt{n}\theta_{2,n}} \times z_{\alpha/2}(1 + o_{\mathbf{P}_{f_0}^n}(1)).$$

It can be easily seen that

$$\frac{\theta_{1,n}^2}{n\theta_{2,n}^2} = \frac{\sum_{\nu=1}^{\infty} \frac{F(\varphi_\nu)^2}{1 + \lambda\gamma_\nu + n^{-1}\tau_\nu^2}}{\sum_{\nu=1}^{\infty} \frac{F(\varphi_\nu)^2}{(1 + \lambda\gamma_\nu + n^{-1}\tau_\nu^2)^2}} \geq 1,$$

together with (A.21) we get that

$$\begin{aligned} & \mathbf{P}_{f_0}^n(|F(f_0) - F(\tilde{f}_{n,\lambda})| \leq r_{F,n}(\alpha)) \\ & = \mathbf{P}_{f_0}^n \left(\left| \frac{F(\tilde{f}_{n,\lambda} - f_0 - T_3)}{\sqrt{n}\theta_{2,n}} + \frac{F(T_3)}{\sqrt{n}\theta_{2,n}} \right| \leq \frac{r_{F,n}(\alpha)}{\sqrt{n}\theta_{2,n}} \right) \\ & \geq \mathbf{P}_{f_0}^n \left(\left| \frac{F(\tilde{f}_{n,\lambda} - f_0 - T_3)}{\sqrt{n}\theta_{2,n}} + \frac{F(T_3)}{\sqrt{n}\theta_{2,n}} \right| \leq z_{\alpha/2}(1 + o_{\mathbf{P}_{f_0}^n}(1)) \right) \\ (A.22) \quad & \rightarrow 1 - \alpha. \end{aligned}$$

Notice that when $0 < \sum_{\nu=1}^{\infty} F(\varphi_\nu)^2 < \infty$, $\frac{\theta_{1,n}^2}{n\theta_{2,n}^2} \rightarrow 1$, leading to that the probability in (A.22) approaches exactly $1 - \alpha$. Proof is completed. \square

REFERENCES

- [1] Adams, R. A. (1975). *Sobolev Spaces*. Academic Press, New York-London. Pure and Applied Mathematics, Vol. 65.
- [2] Birkhoff, D. (1908). Boundary Value and Expansion Problems of Ordinary Linear Differential Equations. *Transactions of the American Mathematical Society*, **9**, 373–395.
- [3] Bickel, P. J. and Kleijn, B. J. K. (2012). The semiparametric Bernstein-von Mises theorem. *Annals of Statistics*, **40**, 206–237.
- [4] Bickel, P. J. and Ritov, Y. (2003). Nonparametric estimators which can be “plugged-in”. *Annals of Statistics*, **31**, 1033–1053.
- [5] Courant, R. and Hilbert, D. (1937). *Methods of Mathematical Physics*, Vol. **1**. Interscience Publishers, Inc., New York.
- [6] Cameron, R. H. and Martin, W. T. (1944). Transformations of Wiener Integrals under Translations. *Annals of Mathematics*, **45**, 386–396.
- [7] Castillo, I. (2012). A semiparametric Bernstein-von Mises theorem for Gaussian process priors. *Probability Theory & Related Fields*, **152**, 53–99.
- [8] Castillo, I. (2014). On Bayesian supremum norm contraction rates. *Annals of Statistics*, **42**, 2058–2091.
- [9] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein-von Mises theorem in Gaussian white noise. *Annals of Statistics*, **41**, 1999–2028.
- [10] Castillo, I. and Nickl, R. (2014). On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. *Annals of Statistics*, **42**, 1941–1969.
- [11] Castillo, I. and Rousseau, J. (2015). A Bernstein-von Mises theorem for smooth functionals in semiparametric models. *Annals of Statistics*. In press.
- [12] Cox, D. (1993). An analysis of Bayesian inference for nonparametric regression. *Annals of Statistics*, **21**, 903–923.
- [13] Cheng, G. and Kosorok, M. R. (2008). General frequentist properties of the posterior profile distribution. *Annals of Statistics*, **36**, 1819–1853.
- [14] Cheng, G. and Kosorok, M. R. (2009). The Penalized Profile Sampler. *Journal of Multivariate Analysis*, **100**, 345–362.
- [15] de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory & Related Fields*, **75**, 261–277.
- [16] Eggermont, P. P. B. and LaRiccia, V. N. (2009). *Maximum Penalized Likelihood Estimation, Volume II: Regression*. Springer.
- [17] Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Annals of Statistics*, **27**, 1119–1140.
- [18] Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in L^r -metrics, $1 \leq r \leq \infty$. *Annals of Statistics*, **39**, 2795–3443.
- [19] Ghoshal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics*, **27**, 143–158.
- [20] Ghoshal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, **28**, 500–531.
- [21] Hájek, J. (1962). On linear statistical problems in stochastic processes. *Czechoslovak Mathematical Journal*, **12**, 404–444.
- [22] Hjort, N.L. and Walker, S. G. (2009). Quantile pyramids for Bayesian nonparametrics. *Annals of Statistics*, **37**, 105–131.

- [23] Hoffmann-Jorgensen, Shepp, L. A. and Dudley, R. M. (1979). On the lower tail of Gaussian seminorms. *Annals of Probability*, **7**, 193–384.
- [24] Johnstone, I. M. (2013). *Gaussian Estimation: Sequence and Wavelet Models*.
- [25] Kim, Y. (2006). Bernstein-von Mises theorem for the proportional hazard model. *Annals of Statistics*, **34**, 1678–1700.
- [26] Kim, Y. and Lee, J. (2004). A Bernstein-von Mises theorem in the nonparametric rightcensoring model. *Annals of Statistics*, **32**, 1492–1512.
- [27] Knapik, B. T., van der Vaart, A. W. and van Zanten, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Annals of Statistics*, **39**, 2626–2657.
- [28] Knapik, B. T., Szabó, B. T., van der Vaart, A. W. and van Zanten, J. H. (2013). Bayes procedures for adaptive inference in inverse problems for the white noise model. Preprint.
- [29] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer: New York.
- [30] Kuelbs, J., Li, W. V. and Linde, W. (1994). The Gaussian measure of shited balls. *Probability Theory and Related Fields*, **98**, 143–162.
- [31] Leahu, H. (2011). On the Bernstein-von Mises phenomenon in the Gaussian white noise model. *Electronic Journal of Statistics*, **5**, 373–404.
- [32] Li, W. V. (1999). A Gaussian correlation inequality and its applications to small ball probabilities. *Electronic Communications in Probability*, **4**, 111–118.
- [33] Morris, C. N. (1982). Natural exponential families with quadratic variance functions. *Annals of Statistics*, **10**, 65–80.
- [34] Nickl, R. (2014). Discussion to “Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Annals of Statistics*, to appear.
- [35] Nickl, R. and Szabó, B. A sharp adaptive confidence ball for self-similar functions. Preprint.
- [36] Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, **83**, 1134–1143.
- [37] Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Annals of Probability*, **22**, 1679–1706.
- [38] Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill. New York.
- [39] Rivoirard, V. and Rousseau, J. (2012). Bernstein-von Mises theorem for linear functionals of the density. *Annals of Statistics*, **40**, 1489–1523.
- [40] Shang, Z. (2010). Convergence rate and Bahadur type representation of general smoothing spline M-Estimates. *Electronic Journal of Statistics*, **4**, 1411–1442.
- [41] Shang, Z. and Cheng, G. (2013). Local and global asymptotic inference in smoothing spline models. *Annals of Statistics*, **41**, 2608–2638.
- [42] Shang, Z. and Cheng, G. (2015). Supplementary document to “Nonparametric Bernstein-von Mises Phenomenon: A Tuning Prior Perspective.”
- [43] Shepp, L. A. (1966). Radon-Nikodym derivatives of Gaussian measures. *Annals of Mathematical Statistics*, **37**, 321–354.
- [44] Sniekers, S. and van der Vaart, A. W. (2014). Credible sets in the fixed design model with Brownian motion prior. *Journal of Statistical Planning and Inference*. In press.
- [45] Sniekers, S. and van der Vaart, A. W. Adaptive Bayesian credible sets in regression with a Gaussian process prior. Preprint.
- [46] Szabó, B. On Bayesian based adaptive confidence sets for linear functionals. Preprint.
- [47] Szabó, B., van der Vaart, A. W. and van Zanten, H. (2014). Empirical Bayes scaling of Gaussian priors in the white noise model. *Electronic Journal of Statistics*, **7**, 991–1018.

- [48] Szabó, B., van der Vaart, A. W. and van Zanten, H. (2014). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Annals of Statistics*.
- [49] Tusnady, G. (1977). A remark on the approximation of the sample distribution function in the multidimensional case. *Periodica Mathematica Hungarica*, **8**, 53–55.
- [50] van der Geer, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge University Press, New York.
- [51] van der Vaart, A. W. and van Zanten, J. H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electronic Journal of Statistics*, **1**, 433–448.
- [52] van der Vaart, A. W. and van Zanten, J. H. (2008). Reproducing kernel Hilbert spaces of Gaussian priors. *IMS Collections. Publishing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*. **3**, 200–222.
- [53] van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Annals of Statistics*, **36**, 1031–1508.
- [54] van der Vaart, A. W. and van Zanten, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. *Annals of Statistics*, **37**, 2655–2675.
- [55] Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Annals of Statistics*, **13**, 1251–1638.
- [56] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- [57] Weinberger, H. F. (1974). *Variational Methods for Eigenvalue Approximation*. Society for Industrial and Applied Mathematics, Philadelphia, Pa.
- [58] Yoo, W. and Ghoshal, S. Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. Preprint.
- [59] Zhao, L. H. (2000). Bayesian aspects of some nonparametric problems. *Annals of Statistics*, **28**, 532–552.

Supplementary document to

NONPARAMETRIC BERNSTEIN-VON MISES PHENOMENON: A TUNING PRIOR PERSPECTIVE

BY ZUOFENG SHANG[‡] AND GUANG CHENG[§]

*Department of Statistics
Purdue University*

This supplementary document contains the proofs of several additional results not included in the main manuscript.

We organize this document as follows:

- Section S.1 contains some preliminary results including the theoretical foundations of the main results.
- Section S.2 includes the proof of Lemma 3.1.
- Section S.3 includes the proof of Theorem 4.1.
- Section S.4 includes the proof of Theorem 5.1.
- Section S.5 addresses the issue of L^2 -diameters of $R_n^\omega(\alpha)$ and $R_n^{\star\omega}(\alpha)$.
- Section S.6 includes the proof of Proposition 6.5.

[‡]Visiting Assistant Professor

[§]Corresponding Author. Associate Professor, Research Sponsored by NSF CAREER Award DMS-1151692, DMS-1418042, Simons Foundation 305266. Guang Cheng is on sabbatical at Princeton while this work is finalized; he would like to thank the Princeton ORFE department for its hospitality and support.

S.1. Some Preliminary Results. In this section, let us introduce some technical preliminaries. Using (5.5), for any $g = \sum_{\nu} g_{\nu} \varphi_{\nu}$, $\tilde{g} = \sum_{\nu} \tilde{g}_{\nu} \varphi_{\nu} \in S^m(\mathbb{I})$, we have $J(g, \tilde{g}) = \sum_{\nu \geq 1} g_{\nu} \tilde{g}_{\nu} \gamma_{\nu}$. It therefore holds that

$$(S.1) \quad J(\varphi_{\nu}, \varphi_{\mu}) = \gamma_{\nu} \delta_{\nu\mu}, \quad \nu, \mu \geq 1.$$

This shows that

$$\|g\|_{V,U}^2 = \sum_{\nu \geq 1} g_{\nu}^2 (1 + \rho_{\nu}), \quad J(g) = \sum_{\nu \geq 1} g_{\nu}^2 \gamma_{\nu}.$$

Since $\gamma_{\nu} \asymp 1 + \rho_{\nu}$, we can see that the $\|\cdot\|_{V,U}$ -norm and $J^{1/2}$ -norm are equivalent. By Sobolev embedding theorem ([1]) which implies that the supremum norm is “weaker” than the $\|\cdot\|_{V,U}$ -norm, there exists an absolute constant $C_3 > 0$ s.t. for any $g \in S^m(\mathbb{I})$,

$$(S.2) \quad \|g\|_{\infty} \leq C_3 \sqrt{J(g)}.$$

For any $f, g, \tilde{g} \in S^m(\mathbb{I})$, define $V_f(g, \tilde{g}) = E\{\ddot{A}(f(X))g(X)\tilde{g}(X)\}$. In particular, $V_{f_0}(\cdot, \cdot) = V(\cdot, \cdot)$. Let $(\varphi_{f,\nu}, \rho_{f,\nu})$ be the eigen-system corresponding to the following ODE:

$$(S.3) \quad \begin{aligned} (-1)^m \varphi_{f,\nu}^{(2m)}(\cdot) &= \rho_{f,\nu} \ddot{A}(f(\cdot)) \pi(\cdot) \varphi_{f,\nu}(\cdot), \\ \varphi_{f,\nu}^{(j)}(0) &= \varphi_{f,\nu}^{(j)}(1) = 0, \quad j = m, m+1, \dots, 2m-1. \end{aligned}$$

It follows from [41, Proposition 2.2] that $(\varphi_{f,\nu}, \rho_{f,\nu})$ satisfy the properties stated in Proposition 2.1 with V therein replaced by V_f . Let $\gamma_{f,\nu} = 1$ if $\nu = 1, 2, \dots, m$; $= \rho_{f,\nu}$ if $\nu > m$. For any $g, \tilde{g} \in S^m(\mathbb{I})$ with $g = \sum_{\nu} g_{\nu} \varphi_{f,\nu}$ and $\tilde{g} = \sum_{\nu} \tilde{g}_{\nu} \varphi_{f,\nu}$, define $J_f(g, \tilde{g}) = \sum_{\nu} g_{\nu} \tilde{g}_{\nu} \gamma_{f,\nu}$. Define an inner product

$$\langle g, \tilde{g} \rangle_f = V_f(g, \tilde{g}) + \lambda J_f(g, \tilde{g}), \quad g \in S^m(\mathbb{I}),$$

and let $\|\cdot\|_f$ be the corresponding norm. Let \mathcal{P}_{λ}^f be a self-adjoint positive-definite operator from $S^m(\mathbb{I})$ to itself s.t. $\langle \mathcal{P}_{\lambda}^f g, \tilde{g} \rangle_f = \lambda J_f(g, \tilde{g})$ for any $g, \tilde{g} \in S^m(\mathbb{I})$. For convenience, define $\mathcal{P}_{\lambda} = \mathcal{P}_{\lambda}^{f_0}$. In particular,

$$J_{f_0}(g, \tilde{g}) = J(g, \tilde{g}), \quad \langle g, \tilde{g} \rangle_{f_0} = \langle g, \tilde{g} \rangle, \quad \|g\|_{f_0} = \|g\|.$$

For any constant C with $C > \|f_0\|_{\infty}$, let C_0, C_1, C_2 be positive constants satisfying Assumption A1. Since $1/C_2 \leq \ddot{A}(z) \leq C_2$ if $|z| \leq 2C$ (Assumption A1), we get that for any $f \in \mathcal{F}(C)$ and $g \in S^m(\mathbb{I})$, (leading to that $C_2^{-1} \leq \ddot{A}(f(X)) \leq C_2$ a.s.)

$$(S.4) \quad C_2^{-2} V(g, g) \leq V_f(g, g) \leq C_2^2 V(g, g),$$

that is, V_f is *uniformly equivalent* to V for $f \in \mathcal{F}(C)$. This leads to

$$C_2^{-2} \frac{V(g)}{V(g) + U(g)} \leq \frac{V_f(g)}{V_f(g) + U(g)} \leq C_2^2 \frac{V(g)}{V(g) + U(g)}.$$

It follows from (S.4) and mapping principle (see [57, Theorem 5.3]) that

$$C_2^{-2} \rho_{\nu} \leq \rho_{f,\nu} \leq C_2^2 \rho_{\nu}, \quad \text{for any } \nu > m \text{ and } f \in \mathcal{F}(C).$$

The following lemma says that the norms $\|\cdot\|$ and $\|\cdot\|_f$ are equivalent.

Lemma S.1. *If $0 < \lambda \leq \frac{1}{2C_2^2}$, then for any $f \in \mathcal{F}(C)$ and $g \in S^m(\mathbb{I})$,*

$$\frac{1}{\sqrt{2}C_2} \|g\| \leq \|g\|_f \leq \sqrt{2}C_2 \|g\|,$$

$$\left(1 + \frac{C_2^2}{\rho_{m+1}}\right)^{-1} C_2^{-2} J(g) \leq J_f(g) \leq \left(1 + \frac{1}{\rho_{m+1}}\right) C_2^2 J(g).$$

PROOF OF LEMMA S.1. For any $g \in S^m(\mathbb{I})$ with $g = \sum_{\nu} g_{\nu} \varphi_{f,\nu}$, we have

$$V_f(g) = \sum_{\nu \geq 1} g_{\nu}^2, \quad U(g) = \sum_{\nu > m} g_{\nu}^2 \rho_{f,\nu}, \quad J_f(g) = \sum_{\nu=1}^m g_{\nu}^2 + \sum_{\nu > m} g_{\nu}^2 \rho_{f,\nu}.$$

So, $J_f(g) \leq V_f(g) + U(g)$ and $U(g) \leq J_f(g)$. Therefore, it follows by (S.4) that

$$\begin{aligned} \|g\|_f^2 &= V_f(g) + \lambda J_f(g) \\ &\leq (1 + \lambda) V_f(g) + \lambda U(g) \\ &\leq (1 + \lambda) C_2^2 V(g) + \lambda J(g) \leq (1 + \lambda) C_2^2 (V(g) + \lambda J(g)) \leq 2C_2^2 \|g\|^2, \end{aligned}$$

where the last inequality is because $\lambda \leq \frac{1}{2C_2^2} < 1$.

On the other hand,

$$\begin{aligned} \|g\|_f^2 &= V_f(g) + \lambda J_f(g) \\ &\geq C_2^{-2} V(g) + \lambda U(g) \\ &\geq C_2^{-2} V(g) + \lambda (J(g) - V(g)) \\ &= (C_2^{-2} - \lambda) V(g) + \lambda J(g) \geq \frac{1}{2C_2^2} (V(g) + \lambda J(g)) = \frac{1}{2C_2^2} \|g\|^2. \end{aligned}$$

Meanwhile, $J_f(g) \leq V_f(g) + U(g) \leq C_2^2 V(g) + J(g)$. It can be shown that $V(g) + U(g) \leq (1 + 1/\rho_{m+1}) J(g)$. To see this, write $g = \sum_{\nu} g_{\nu} \varphi_{\nu}$. Then it follows by $1 + \rho_{\nu} \leq (1 + 1/\rho_{m+1}) \gamma_{\nu}$ that

$$V(g) + U(g) = \sum_{\nu} g_{\nu}^2 (1 + \rho_{\nu}) \leq (1 + 1/\rho_{m+1}) \sum_{\nu} g_{\nu}^2 \gamma_{\nu} = (1 + 1/\rho_{m+1}) J(g).$$

So $J_f(g) \leq (1 + 1/\rho_{m+1}) C_2^2 J(g)$.

Similarly, we have that $J(g) \leq V(g) + U(g) \leq C_2^2 V_f(g) + U(g)$. Write $g = \sum_{\nu} g_{\nu} \varphi_{f,\nu}$. Since $C_2^2 \rho_{\nu} \geq \rho_{f,\nu} \geq C_2^{-2} \rho_{\nu} \geq C_2^{-2} \rho_{m+1}$ for $\nu > m$, we have $1 + \rho_{f,\nu} \leq (1 + C_2^2/\rho_{m+1}) \gamma_{f,\nu}$. So

$$\begin{aligned} V_f(g) + U(g) &= \sum_{\nu} g_{\nu}^2 (1 + \rho_{f,\nu}) \\ &\leq (1 + C_2^2/\rho_{m+1}) \sum_{\nu} g_{\nu}^2 \gamma_{f,\nu} = (1 + C_2^2/\rho_{m+1}) J_f(g). \end{aligned}$$

Therefore, $J_f(g) \geq (1 + C_2^2/\rho_{m+1})^{-1} C_2^{-2} J(g)$. Proof is completed. \square

The equivalence of $\|\cdot\|$ and $\|\cdot\|_f$ stated in Lemma S.1 leads to that $S^m(\mathbb{I})$ is a RKHS under $\langle \cdot, \cdot \rangle_f$ for any $f \in \mathcal{F}(C)$. Let $K^f(x, x')$ be the corresponding reproducing kernel function. In particular, $K^{f_0} = K$. By [41, Proposition 2.1] we have the following series representation.

Proposition S.1. *For any $f \in \mathcal{F}(C)$, $g \in S^m(\mathbb{I})$ and $x \in \mathbb{I}$, we have $\|g\|_f^2 = \sum_{\nu} |V_f(g, \varphi_{f,\nu})|^2 (1 + \lambda \gamma_{f,\nu})$, $K_x^f(\cdot) \equiv K^f(x, \cdot) = \sum_{\nu} \frac{\varphi_{f,\nu}(x)}{1 + \lambda \gamma_{f,\nu}} \varphi_{f,\nu}(\cdot)$, and $\mathcal{P}_{\lambda}^f \varphi_{f,\nu}(\cdot) = \frac{\lambda \gamma_{f,\nu}}{1 + \lambda \gamma_{f,\nu}} \varphi_{f,\nu}(\cdot)$.*

The following lemma demonstrates a uniform bound for the kernel K^f .

Lemma S.2. *It holds that*

$$c_K(C) \equiv \sup_{f \in \mathcal{F}(C)} \sup_{0 < h \leq 1} \sup_{x \in \mathbb{I}} h^{1/2} \|K_x^f\|_f \leq c_m \sqrt{\frac{C_2}{\pi}} + 1,$$

where $c_m > 0$ is a universal constant depending on m only.

PROOF OF LEMMA S.2. For any $f \in \mathcal{F}(C)$, $g \in S^m(\mathbb{I})$ and $x \in \mathbb{I}$, it follows by [16, Lemma (2.11), pp. 54] that

$$|\langle K_x^f, g \rangle_f| = |g(x)| \leq c_m h^{-1/2} \sqrt{\|g\|_{L^2}^2 + \lambda \|g^{(m)}\|_{L^2}^2},$$

where $c_m > 0$ is a universal constant depending on m only, and $\|\cdot\|_{L^2}$ denotes the usual L^2 -norm. Since $\|g\|_{L^2}^2 \leq \frac{C_2}{\pi} V_f(g)$ and $\|g^{(m)}\|_{L^2}^2 = U(g) \leq J_f(g)$ (see proof of Lemma S.1 for the last inequality). Then

$$|\langle K_x^f, g \rangle_f| \leq c_m \sqrt{\frac{C_2}{\pi}} + 1 h^{-1/2} \|g\|_f,$$

implying that $\|K_x^f\|_f \leq c_m \sqrt{\frac{C_2}{\pi}} + 1 h^{-1/2}$. So $c_K(C) \leq c_m \sqrt{\frac{C_2}{\pi}} + 1$. \square

The lemma below directly comes from Lemma S.2, which relates the norms $\|\cdot\|_f$ and $\|\cdot\|_{\infty}$.

Lemma S.3. *For any $f \in \mathcal{F}(C)$ and $g \in S^m(\mathbb{I})$, $\|g\|_{\infty} \leq c_K(C) h^{-1/2} \|g\|_f$.*

Suppose that (Y, X) follows model (2.1) based on f . The following conditional expectation can be found based on [33]:

$$(S.5) \quad E_f\{Y|X\} = \dot{A}(f(X)).$$

Let $g, g_k \in S^m(\mathbb{I})$ for $k = 1, 2, 3$. The Fréchet derivative of $\ell_{n,\lambda}$ can be identified as

$$\begin{aligned} D\ell_{n,\lambda}(g)g_1 &= \frac{1}{n} \sum_{i=1}^n (Y_i - \dot{A}(g(X_i))) \langle K_{X_i}^f, g_1 \rangle_f - \langle \mathcal{P}_{\lambda}^f g, g_1 \rangle_f \\ &\equiv \langle S_{n,\lambda}(g), g_1 \rangle_f. \end{aligned}$$

Define $S_\lambda(g) = E_f\{S_{n,\lambda}(g)\}$. We also use DS_λ and D^2S_λ to represent the second- and third-order Fréchet derivatives of S_λ . Note that $S_{n,\lambda}(\hat{f}_{n,\lambda}) = 0$, and $S_{n,\lambda}(f)$ can be expressed as

$$(S.6) \quad S_{n,\lambda}(f) = \frac{1}{n} \sum_{i=1}^n (Y_i - \dot{A}(f(X_i))) K_{X_i}^f - \mathcal{P}_\lambda^f f.$$

The Fréchet derivatives of $S_{n,\lambda}$ and $DS_{n,\lambda}$ are denoted $DS_{n,\lambda}(g)g_1g_2$ and $D^2S_{n,\lambda}(g)g_1g_2g_3$. These derivatives can be explicitly written as

$$\begin{aligned} D^2\ell_{n,\lambda}(g)g_1g_2 &\equiv DS_{n,\lambda}(g)g_1g_2 \\ &= -\frac{1}{n} \sum_{i=1}^n \ddot{A}(g(X_i))g_1(X_i)g_2(X_i) - \langle \mathcal{P}_\lambda^f g_1, g_2 \rangle_f, \\ D^3\ell_{n,\lambda}(g)g_1g_2g_3 &\equiv D^2S_{n,\lambda}(g)g_1g_2g_3 \\ &= -\frac{1}{n} \sum_{i=1}^n \ddot{A}(g(X_i))g_1(X_i)g_2(X_i)g_3(X_i), \\ DS_\lambda(g)g_1 &= -E\{\ddot{A}(g(X))g_1(X)K_X^f\} - \mathcal{P}_\lambda^f g_1, \\ D^2S_\lambda(g)g_1g_2 &= -E\{\ddot{A}(g(X))g_1(X)g_2(X)K_X^f\}. \end{aligned}$$

Consider a function class

$$\mathcal{G}(C) = \{g \in S^m(\mathbb{I}) : \|g\|_\infty \leq 1, J(g, g) \leq 2C_2^2 c_K(C)^{-2} h^{-2m+1}\}.$$

Let $N(\varepsilon, \mathcal{G}(C), \|\cdot\|_\infty)$ be ε packing number in terms of supremum norm. The following result can be found in [50].

Lemma S.4. *There exists a universal constant $c_0 > 0$ s.t. for any $\varepsilon > 0$,*

$$\log N(\varepsilon, \mathcal{G}(C), \|\cdot\|_\infty) \leq c_0 (\sqrt{2}C_2 c_K(C)^{-1})^{1/m} h^{-\frac{2m-1}{2m}} \varepsilon^{-1/m}.$$

In the future, for notational simplicity, we will simply drop C from $c_K(C)$ and $\mathcal{G}(C)$ if there is no confusion.

For $r \geq 0$, define $\Psi(r) = \int_0^r \sqrt{\log(1 + \exp(x^{-1/m}))} dx$. For arbitrary $\varepsilon > 0$, define

$$\begin{aligned} A(h, \varepsilon) &= \frac{32\sqrt{6}}{\tau} \sqrt{2}C_2 c_K^{-1} c_0^m h^{-(2m-1)/2} \Psi\left(\frac{1}{2\sqrt{2}C_2} c_K c_0^{-m} h^{(2m-1)/2} \varepsilon\right) \\ &\quad + \frac{10\sqrt{24}\varepsilon}{\tau} \sqrt{\log\left(1 + \exp\left(2c_0((\sqrt{2}C_2)^{-1} c_K h^{(2m-1)/2} \varepsilon)^{-1/m}\right)\right)}, \end{aligned}$$

where $\tau = \sqrt{\log 1.5} \approx 0.6368$. We have the following useful lemma.

Lemma S.5. *For any $f \in S^m(\mathbb{I})$, suppose that $\psi_{n,f}(z; g)$ is a measurable function defined upon $z = (y, x) \in \mathcal{Y} \times \mathbb{I}$ and $g \in \mathcal{G}$ satisfying $\psi_{n,f}(z; 0) = 0$ and the following Lipschitz continuity condition: for any $1 \leq i \leq n$ and $g_1, g_2 \in \mathcal{G}$,*

$$(S.7) \quad |\psi_{n,f}(Z_i; g_1) - \psi_{n,f}(Z_i; g_2)| \leq c_K^{-1} h^{1/2} \|g_1 - g_2\|_\infty.$$

Then for any constant $t \geq 0$ and $n \geq 1$,

$$\sup_{f \in S^m(\mathbb{I})} \mathbb{P}_f \left(\sup_{g \in \mathcal{G}} \|Z_{n,f}(g)\|_f > t \right) \leq 2 \exp \left(-\frac{t^2}{B(h)^2} \right),$$

where $B(h) = A(h, 2)$ and

$$Z_{n,f}(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\psi_{n,f}(Z_i; g) K_{X_i}^f - E_f \{ \psi_{n,f}(Z_i; g) K_{X_i}^f \}].$$

PROOF OF LEMMA S.5. For any $f \in S^m(\mathbb{I})$ and $n \geq 1$, and any $g_1, g_2 \in \mathcal{G}$, we get that

$$\begin{aligned} & \|(\psi_{n,f}(Z_i; g_1) - \psi_{n,f}(Z_i; g_2)) K_{X_i}^f\|_f \\ & \leq c_K^{-1} h^{1/2} \|g_1 - g_2\|_\infty c_K h^{-1/2} = \|g_1 - g_2\|_\infty. \end{aligned}$$

By Theorem 3.5 of [37], for any $t > 0$, $\mathbb{P}_f (\|Z_{n,f}(g_1) - Z_{n,f}(g_2)\|_f \geq t) \leq 2 \exp \left(-\frac{t^2}{8\|g_1 - g_2\|_\infty^2} \right)$. Then by Lemma 8.1 in [29], we have

$$\| \|Z_{n,f}(g_1) - Z_{n,f}(g_2)\|_f \|_{\psi_2} \leq \sqrt{24} \|g_1 - g_2\|_\infty,$$

where $\|\cdot\|_{\psi_2}$ denotes the Orlicz norm associated with $\psi_2(s) := \exp(s^2) - 1$. Recall $\tau = \sqrt{\log 1.5} \approx 0.6368$. Define $\phi(x) = \psi_2(\tau x)$. Then it can be shown by elementary calculus that $\phi(1) \leq 1/2$, and for any $x, y \geq 1$, $\phi(x)\phi(y) \leq \phi(xy)$. By a careful examination of the proof of Lemma 8.2, it can be shown that for any random variables ξ_1, \dots, ξ_l ,

$$(S.8) \quad \left\| \max_{1 \leq i \leq l} \xi_i \right\|_{\psi_2} \leq \frac{2}{\tau} \psi_2^{-1}(l) \max_{1 \leq i \leq l} \|\xi_i\|_{\psi_2}.$$

Next we use a “chaining” argument. Let $T_0 \subset T_1 \subset T_2 \subset \dots \subset T_\infty := \mathcal{G}$ be a sequence of finite nested sets satisfying the following properties:

- for any T_q and any $s, t \in T_q$, $\|s - t\|_\infty \geq \varepsilon 2^{-q}$; each T_q is “maximal” in the sense that if one adds any point in T_q , then the inequality will fail;
- the cardinality of T_q is upper bounded by

$$\begin{aligned} \log |T_q| & \leq \log N(\varepsilon 2^{-q}, \mathcal{G}, \|\cdot\|_\infty) \\ & \leq c_0 (\sqrt{2} C_2 c_K^{-1})^{1/m} h^{-(2m-1)/(2m)} (\varepsilon 2^{-q})^{-1/m}, \end{aligned}$$

where $c_0 > 0$ is absolute constant;

- each element $t_{q+1} \in T_{q+1}$ is uniquely linked to an element $t_q \in T_q$ which satisfies $\|t_q - t_{q+1}\|_\infty \leq \varepsilon 2^{-q}$.

For arbitrary $s_{k+1}, t_{k+1} \in T_{k+1}$ with $\|s_{k+1} - t_{k+1}\|_\infty \leq \varepsilon$, choose two chains (both being of length $k+2$) t_q and s_q with $t_q, s_q \in T_q$ for $0 \leq q \leq k+1$. The ending points s_0 and t_0 satisfy

$$\begin{aligned} \|s_0 - t_0\|_\infty & \leq \sum_{q=0}^k [\|s_q - s_{q+1}\|_\infty + \|t_q - t_{q+1}\|_\infty] + \|s_{k+1} - t_{k+1}\|_\infty \\ & \leq 2 \sum_{q=0}^k \varepsilon 2^{-q} + \varepsilon \leq 5\varepsilon, \end{aligned}$$

and hence, $\|Z_{n,f}(s_0) - Z_{n,f}(t_0)\|_f\|_{\psi_2} \leq 5\sqrt{24}\varepsilon$. It follows by the proof of Theorem 8.4 of [29] and (S.8) that

$$\begin{aligned}
& \left\| \max_{s_{k+1}, t_{k+1} \in T_{k+1}} \|Z_{n,f}(s_{k+1}) - Z_{n,f}(t_{k+1}) - (Z_{n,f}(s_0) - Z_{n,f}(t_0))\|_f \right\|_{\psi_2} \\
& \leq 2 \sum_{q=0}^k \left\| \max_{\substack{u \in T_{q+1}, v \in T_q \\ u, v \text{ link each other}}} \|Z_{n,f}(u) - Z_{n,f}(v)\|_f \right\|_{\psi_2} \\
& \leq \frac{4}{\tau} \sum_{q=0}^k \psi_2^{-1}(N(2^{-q-1}\varepsilon, \mathcal{G}, \|\cdot\|_\infty)) \\
& \quad \times \max_{\substack{u \in T_{q+1}, v \in T_q \\ u, v \text{ link each other}}} \|Z_{n,f}(u) - Z_{n,f}(v)\|_f\|_{\psi_2} \\
& \leq \frac{4\sqrt{24}}{\tau} \sum_{q=0}^k \sqrt{\log(1 + N(\varepsilon 2^{-q-1}, \mathcal{G}, \|\cdot\|_\infty))} \varepsilon 2^{-q} \\
& \leq \frac{8\sqrt{24}}{\tau} \sum_{q=1}^{k+1} \sqrt{\log\left(1 + \exp\left(c_0 c_K^{-1/m} h^{-(2m-1)/(2m)} (\varepsilon 2^{-q})^{-1/m}\right)\right)} \varepsilon 2^{-q} \\
& \leq \frac{32\sqrt{6}}{\tau} \int_0^{\varepsilon/2} \sqrt{\log\left(1 + \exp\left(c_0 c_K^{-1/m} h^{-(2m-1)/(2m)} x^{-1/m}\right)\right)} dx \\
& = \frac{32\sqrt{6}}{\tau} c_K^{-1} c_0^m h^{-(2m-1)/2} \Psi\left(\frac{1}{2} c_K c_0^{-m} h^{(2m-1)/2} \varepsilon\right).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& \left\| \max_{\substack{u, v \in T_0 \\ \|u-v\|_\infty \leq 5\varepsilon}} \|Z_{n,f}(u) - Z_{n,f}(v)\|_f \right\|_{\psi_2} \\
& \leq \frac{2}{\tau} \psi_2(|T_0|^2) \max_{\substack{u, v \in T_0 \\ \|u-v\|_\infty \leq 5\varepsilon}} \|Z_{n,f}(u) - Z_{n,f}(v)\|_f\|_{\psi_2} \\
& \leq \frac{2}{\tau} \psi_2^{-1}(N(\varepsilon, \mathcal{G}, \|\cdot\|_\infty)^2) (5\sqrt{24}\varepsilon).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \left\| \max_{\substack{s, t \in T_{k+1} \\ \|s-t\|_\infty \leq \varepsilon}} \|Z_{n,f}(s) - Z_{n,f}(t)\|_f \right\|_{\psi_2} \\
& \leq \frac{32\sqrt{6}}{\tau} c_K^{-1} c_0^m h^{-(2m-1)/2} \Psi \left(\frac{1}{2} c_K c_0^{-m} h^{(2m-1)/2} \varepsilon \right) \\
& \quad + \frac{2}{\tau} \psi_2^{-1} (N(\varepsilon, \mathcal{G}, \|\cdot\|_\infty)^2) (5\sqrt{24}\varepsilon) \\
& \leq \frac{32\sqrt{6}}{\tau} c_K^{-1} c_0^m h^{-(2m-1)/2} \Psi \left(\frac{1}{2} c_K c_0^{-m} h^{(2m-1)/2} \varepsilon \right) \\
& \quad + \frac{10\sqrt{24}\varepsilon}{\tau} \sqrt{\log(1 + \exp(2c_0(c_K h^{(2m-1)/2} \varepsilon)^{-1/m}))} \\
& = A(h, \varepsilon).
\end{aligned}$$

Now for any $g_1, g_2 \in \mathcal{G}$ with $\|g_1 - g_2\|_\infty \leq \varepsilon/2$. Let $k \geq 2$, hence, $2^{1-k} \leq 1 - \|g_1 - g_2\|_\infty/\varepsilon$. Since T_k is “maximal”, there exist $s_k, t_k \in T_k$ s.t. $\max\{\|g_1 - s_k\|_\infty, \|g_2 - t_k\|_\infty\} \leq \varepsilon 2^{-k}$. It is easy to see that $\|s_k - t_k\|_\infty \leq \varepsilon$. So

$$\begin{aligned}
& \|Z_{n,f}(g_1) - Z_{n,f}(g_2)\|_f \\
& \leq \|Z_{n,f}(g_1) - Z_{n,f}(s_k)\|_f + \|Z_{n,f}(g_2) - Z_{n,f}(t_k)\|_f \\
& \quad + \|Z_{n,f}(s_k) - Z_{n,f}(t_k)\|_f \\
& \leq 4\sqrt{n}\varepsilon 2^{-k} + \max_{\substack{u, v \in T_k \\ \|u-v\|_\infty \leq \varepsilon}} \|Z_{n,f}(u) - Z_{n,f}(v)\|_f.
\end{aligned}$$

Therefore, letting $k \rightarrow \infty$ we get that

$$\begin{aligned}
& \left\| \sup_{\substack{g_1, g_2 \in \mathcal{G} \\ \|g_1 - g_2\|_\infty \leq \varepsilon/2}} \|Z_{n,f}(g_1) - Z_{n,f}(g_2)\|_f \right\|_{\psi_2} \\
& \leq 4\sqrt{n}\varepsilon 2^{-k} / \sqrt{\log 2} + \left\| \max_{\substack{u, v \in T_k \\ \|u-v\|_\infty \leq \varepsilon}} \|Z_{n,f}(u) - Z_{n,f}(v)\|_f \right\|_{\psi_2} \\
& \leq 4\sqrt{n}\varepsilon 2^{-k} / \sqrt{\log 2} + A(h, \varepsilon) \rightarrow A(h, \varepsilon).
\end{aligned}$$

Taking $\varepsilon = 2$ in the above inequality, we get that

$$\left\| \sup_{\substack{g_1, g_2 \in \mathcal{G} \\ \|g_1 - g_2\|_\infty \leq 1}} \|Z_{n,f}(g_1) - Z_{n,f}(g_2)\|_f \right\|_{\psi_2} \leq A(h, 2) = B(h).$$

By Lemma 8.1 in [29], we have

$$\mathbb{P}_f \left(\sup_{g \in \mathcal{G}} \|Z_{n,f}(g)\|_f \geq t \right) \leq 2 \exp \left(-\frac{t^2}{B(h)^2} \right).$$

Note that the right hand side in the above does not depend on f . This completes the proof. \square

Define

$$H^m(C) = \{f \in S^m(\mathbb{I}) : J(f) \leq C^2/C_3^2\}.$$

It follows from (S.2) that for any $g \in H^m(C)$, $\|g\|_\infty \leq C_3\sqrt{J(g)} \leq C$, implying that $g \in \mathcal{F}(C)$. Thus, we have proved the following inclusion:

$$(S.9) \quad H^m(C) \subseteq \mathcal{F}(C).$$

It is easy to see that when $C > C_3\sqrt{J(f_0)}$, then $f_0 \in H^m(C)$, and hence, $f_0 \in \mathcal{F}(C)$.

Lemma S.6. *Suppose that Assumption A1 holds. For any constant C satisfying $C > C_3\sqrt{J(f_0)}$, let C_0, C_1, C_2 be positive constants satisfying Assumption A1, and define*

$$(S.10) \quad b = \frac{C_2 C}{C_3} \sqrt{1 + \frac{1}{\rho_{m+1}}}.$$

If r, h, M are positives satisfying the following Rate Condition (**H**):

- (i) $(4C_2c_K^2 + 5)bh^{m-1/2} \leq 2(\log 2)C_0c_K$, $C_2^2c_Kbh^{m-1/2} \leq 1/4$,
 $2bC_2h^{m+1/2} \leq c_K$,
- (ii) $h^{1/2}r \leq 1$,
- (iii) $C_2c_K^2M^{1/2}rh^{-1/2}B(h) \leq 1/6$,
- (iv) $12C_0C_2c_K^4(4C_1 + M)h^{-1}r(M^{1/2}rB(h) + C_2^{1/2}c_K^{-1}) \leq 1/6$,

then, for any $1 \leq j \leq s$, the following two results hold:

(a)

$$\sup_{f \in H^m(C)} \mathbf{P}_f \left(\|\hat{f}_{n,\lambda} - f\|_f \geq \delta_n \right) \leq 6 \exp(-Mnhr^2),$$

where $\delta_n = 2bh^m + 24C_0c_K(4C_1 + M)r$;

(b) if in addition, $c_Kh^{-1/2}\delta_n < C$, then

$$\sup_{f \in H^m(C)} \mathbf{P}_f \left(\|\hat{f}_{n,\lambda} - f - S_{n,\lambda}(f)\|_f > a_n + b_n \right) \leq 8 \exp(-Mnhr^2),$$

where

$$a_n = C_2c_K^2M^{1/2}h^{-1/2}rB(h)\delta_n, \text{ and } b_n = C_2^2c_Kh^{-1/2}\delta_n^2.$$

We remark that Part (b) of Lemma S.6 can be viewed as a uniform extension of the functional Bahadur representation established by [40, 41].

PROOF OF LEMMA S.6. Let $f \in H^m(C)$ be the parameter based on which the data are drawn. It is easy to see that

$$DS_\lambda(f)g = -E\{A(f(X))g(X)K_X^f\} - \mathcal{P}_\lambda^f g,$$

for any $g \in S^m(\mathbb{I})$. Therefore, for any $g, \tilde{g} \in S^m(\mathbb{I})$, $\langle DS_\lambda(f)g, \tilde{g} \rangle_f = -\langle g, \tilde{g} \rangle_f$, leading to $DS_\lambda(f) = -id$.

The proof of (a) is finished in two parts.

Part I: Define an operator mapping $S^m(\mathbb{I})$ to $S^m(\mathbb{I})$:

$$T_{1f}(g) = g + S_\lambda(f + g), \quad g \in S^m(\mathbb{I}).$$

First observe that

$$\|S_\lambda(f)\|_f = \|\mathcal{P}_\lambda^f f\|_f = \sup_{\|g\|_f=1} |\langle \mathcal{P}_\lambda^f f, g \rangle_f| \leq \sqrt{\lambda J_f(f)} \leq h^m b,$$

where the last inequality follows by Lemma S.1 and $f \in H^m(C)$. Let $r_{1n} = 2bh^m$. Let $\mathbb{B}(r_{1n}) = \{g \in S^m(\mathbb{I}) : \|g\|_f \leq r_{1n}\}$ be the r_{1n} -ball. For any $g \in \mathbb{B}(r_{1n})$, using $DS_\lambda(f) = -id$ and $\|g\|_\infty \leq c_K h^{-1/2} r_{1n} = 2c_K b h^{m-1/2} \leq C$, it is easy to see that

$$\begin{aligned} & \|T_{1f}(g)\|_f \\ & \leq \|g + S_\lambda(f + g) - S_\lambda(f)\|_f + \|S_\lambda(f)\|_f \\ & = \|g + DS_\lambda(f)g + \int_0^1 \int_0^1 s D^2 S_\lambda(f + ss'g) g g ds ds'\|_f + \|S_\lambda(f)\|_f \\ & = \left\| \int_0^1 \int_0^1 s D^2 S_\lambda(f + ss'g) g g ds ds' \right\|_f + \|S_\lambda(f)\|_f \\ & = \left\| \int_0^1 \int_0^1 s E\{\ddot{A}(f(X) + ss'g(X))g(X)^2 K_X^f\} ds ds' \right\|_f + r_{1n}/2 \\ & \leq C_2 c_K h^{-1/2} \int_0^1 \int_0^1 s E\{g(X)^2\} ds ds' + r_{1n}/2 \\ & \leq C_2^2 c_K h^{-1/2} \|g\|_f^2 / 2 + r_{1n}/2 \\ & \leq C_2^2 c_K h^{-1/2} r_{1n}^2 / 2 + r_{1n}/2 = C_2^2 c_K b h^{m-1/2} r_{1n} + r_{1n}/2 \leq 3r_{1n}/4, \end{aligned}$$

where the last step follows from the assumption $C_2^2 c_K b h^{m-1/2} \leq 1/4$. Therefore, T_{1f} maps $\mathbb{B}(r_{1n})$ to itself.

For any $g_1, g_2 \in \mathbb{B}(r_{1n})$, denote $g = g_1 - g_2$. Note that for any $0 \leq s \leq 1$, $\|g_2 + sg\|_f \leq s\|g_1\|_f + (1-s)\|g_2\|_f \leq r_{1n}$. By rate assumption we get that $\|g_2 + sg\|_\infty \leq c_K h^{-1/2} r_{1n} = 2bc_K h^{m-1/2} < C$, and hence $|f(X) + s'(g_2(X) + sg(X))| \leq 2C$ for any $s, s' \in [0, 1]$. By Taylor's expansion and

Lemma S.4 we have

$$\begin{aligned}
& \|T_{1f}(g_1) - T_{1f}(g_2)\|_f \\
&= \|g_1 - g_2 + S_\lambda(f + g_1) - S_\lambda(f + g_2)\|_f \\
&= \|g_1 - g_2 + \int_0^1 DS_\lambda(f + g_2 + sg)gds\|_f \\
&= \left\| \int_0^1 [DS_\lambda(f + g_2 + sg) - DS_\lambda(f)]gds \right\|_f \\
&= \left\| \int_0^1 \int_0^1 D^2S_\lambda(f + s'(g_2 + sg))(g_2 + sg)gdsds' \right\|_f \\
&\leq \int_0^1 \int_0^1 \|E\{\ddot{A}(f(X) + s'(g_2(X) + sg(X)))(g_2(X) + sg(X))g(X)K_X^f\}\|_f dsds' \\
&\leq C_2 c_K h^{-1/2} \int_0^1 E\{|g_2(X) + sg(X)| \times |g(X)|\} ds \\
&\leq C_2^2 c_K h^{-1/2} \int_0^1 \|g_2 + sg\|_f ds \times \|g\|_f \\
&\leq 2C_2^2 c_K b h^{m-1/2} \|g_1 - g_2\|_f \leq \|g_1 - g_2\|_f / 2.
\end{aligned}$$

This shows that T_{1f} is a contraction mapping which maps $\mathbb{B}(r_{1n})$ into $\mathbb{B}(r_{1n})$. By contraction mapping theorem (see [38]), T_{1f} has a unique fixed point $g' \in \mathbb{B}(r_{1n})$ satisfying $T_{1f}(g') = g'$. Let $f_\lambda = f + g'$. Then $S_\lambda(f_\lambda) = 0$ and $\|f_\lambda - f\|_f \leq r_{1n}$.

Part II: For any $f \in H^m(C)$, under (2.1) with f being the truth, let f_λ be the function obtained in **Part I** s.t. $\|f_\lambda - f\|_f \leq r_{1n}$, and hence, $\|f_\lambda - f\|_\infty \leq c_K h^{-1/2} \|f_\lambda - f\|_f \leq c_K h^{-1/2} r_{1n} \leq C/4$ so that $|f(X) + s(f_\lambda(X) - f(X))| \leq 2C$ a.s. for any $s \in [0, 1]$. It can be shown that for all $g_1, g_2 \in S^m(\mathbb{I})$,

$$\begin{aligned}
& |[DS_\lambda(f_\lambda) - DS_\lambda(f)]g_1g_2| \\
&= \left| \int_0^1 D^2S_\lambda(f + s(f_\lambda - f))(f_\lambda - f)g_1g_2ds \right| \\
&\leq \int_0^1 |E\{\ddot{A}(f(X) + s(f_\lambda - f)(X))(f_\lambda - f)(X)g_1(X)g_2(X)\}| ds \\
&\leq C_2 E\{|f_\lambda(X) - f(X)| \cdot |g_1(X)g_2(X)|\} \\
&\leq 2C_2^2 c_K b h^{m-1/2} \|g_1\|_f \|g_2\|_f \leq \|g_1\|_f \|g_2\|_f / 2.
\end{aligned}$$

where the last inequality follows by $C_2^2 c_K b h^{m-1/2} \leq 1/4$. Together with the fact $DS_\lambda(f) = -id$, we get that the operator norm $\|DS_\lambda(f_\lambda) + id\|_{\text{operator}} \leq 1/2$. This implies that $DS_\lambda(f_\lambda)$ is invertible with operator norm within $[1/2, 3/2]$, and hence, $\|DS_\lambda(f_\lambda)^{-1}\|_{\text{operator}} \leq 2$.

Define an operator

$$T_{2f}(g) = g - [DS_\lambda(f_\lambda)]^{-1} S_{n,\lambda}(f_\lambda + g), \quad g \in S^m(\mathbb{I}).$$

Rewrite T_{2f} as

$$\begin{aligned} T_{2f}(g) &= -DS_\lambda(f_\lambda)^{-1}[DS_{n,\lambda}(f_\lambda)g - DS_\lambda(f_\lambda)g] \\ &\quad -DS_\lambda(f_\lambda)^{-1}[S_{n,\lambda}(f_\lambda + g) - S_{n,\lambda}(f_\lambda) - DS_{n,\lambda}(f_\lambda)g] \\ &\quad -DS_\lambda(f_\lambda)^{-1}S_{n,\lambda}(f_\lambda). \end{aligned}$$

Denote the above three terms by I_{1f}, I_{2f}, I_{3f} , respectively.

For any $1 \leq i \leq n$, let

$$R_i = (Y_i - \dot{A}(f_\lambda(X_i)))K_{X_i}^f - E_f\{(Y - \dot{A}(f_\lambda(X)))K_X^f\}.$$

Since $E_f\{Y - \dot{A}(f(X))|X\} = 0$, it can be shown that for some (random) $s \in [0, 1]$,

$$\begin{aligned} &\|E_f\{(Y - \dot{A}(f_\lambda(X)))K_X^f\}\|_f \\ &= \sup_{\|g\|_f=1} |\langle E_f\{(Y - \dot{A}(f_\lambda(X)))K_X^f\}, g \rangle_f| \\ &= \sup_{\|g\|_f=1} |E_f\{(Y - \dot{A}(f_\lambda(X)))g(X)\}| \\ &= \sup_{\|g\|_f=1} |E_f\{(\dot{A}(f_\lambda(X)) - \dot{A}(f(X)))g(X)\}| \\ &= \sup_{\|g\|_f=1} \left| E_f \left\{ \ddot{A}(f(X))(f_\lambda(X) - f(X))g(X) \right\} \right. \\ &\quad \left. + \frac{1}{2}E_f \left\{ \ddot{A}(f(X) + s(f_\lambda(X) - f(X)))(f_\lambda(X) - f(X))^2g(X) \right\} \right| \\ &= \sup_{\|g\|_f=1} \left| \langle f_\lambda - f, g \rangle_f \right. \\ &\quad \left. + \frac{1}{2}E_f \left\{ \ddot{A}(f(X) + s(f_\lambda(X) - f(X)))(f_\lambda(X) - f(X))^2g(X) \right\} \right| \\ &\leq \|f_\lambda - f\|_f + \frac{C_2}{2}E\{(f_\lambda(X) - f(X))^2|g(X)|\} \\ &\leq \|f_\lambda - f\|_f + \frac{1}{2}C_2^2c_Kh^{-1/2}\|f_\lambda - f\|_f^2 \\ &\leq r_{1n} + C_2^2c_Kbh^{m-1/2}r_{1n} \leq 5r_{1n}/4. \end{aligned}$$

Therefore,

$$\begin{aligned} \|R_i\|_f &\leq c_Kh^{-1/2}|Y_i - \dot{A}(f_\lambda(X_i))| + 5r_{1n}/4 \\ &\leq c_Kh^{-1/2} \left(|Y_i - \dot{A}(f(X_i))| + 2C_2c_Kbh^{m-1/2} \right) + 5r_{1n}/4, \end{aligned}$$

which leads to that

$$E \left\{ \exp \left(\frac{\|R_i\|_f}{C_0c_Kh^{-1/2}} \right) \right\} \leq C_1 \exp \left(\frac{(4C_2^2c_K^2 + 5)bh^{m-1/2}}{2C_0c_K} \right) \leq 2C_1,$$

where the last inequality follows by condition

$$(4C_2c_K^2 + 5)bh^{m-1/2} \leq 2(\log 2)C_0c_K.$$

Let $\delta = hr/(2C_0c_K)$. Recall the condition $h^{1/2}r \leq 1$ which implies $\delta \leq (2C_0c_Kh^{-1/2})^{-1}$. Therefore,

$$E\{\exp(2\delta\|R_i\|_f)\} \leq E\{\exp(\|R_i\|_f/(C_0c_Kh^{-1/2}))\} \leq 2C_1.$$

Moreover, $\|R_i\|_f^2 \leq 8C_0^2c_K^2h^{-1} \exp(\|R_i\|_f/(2C_0c_Kh^{-1/2}))$, which leads to that

$$\begin{aligned} & E\{\exp(\delta\|R_i\|_f) - 1 - \delta\|R_i\|_f\} \\ & \leq E\{(\delta\|R_i\|_f)^2 \exp(\delta\|R_i\|_f)\} \\ & \leq 8C_0^2c_K^2h^{-1}\delta^2 E\left\{\exp\left(\left(\delta + \frac{1}{2C_0c_Kh^{-1/2}}\right)\|R_i\|_f\right)\right\} \\ & \leq 16C_0^2C_1c_K^2h^{-1}\delta^2. \end{aligned}$$

It follows by Theorem 3.2 of [37] that, for $L(M) := 2C_0c_K(4C_1 + M)$,

$$\begin{aligned} \mathbb{P}_f\left(\left\|\sum_{i=1}^n R_i\right\|_f \geq L(M)nr\right) & \leq 2 \exp(-L(M)\delta nr + 16C_0^2C_1c_K^2nh^{-1}\delta^2) \\ (S.11) \qquad \qquad \qquad & = 2 \exp(-Mnhr^2), \end{aligned}$$

We note that the right hand side in the above inequality does not depend on f . It is easy to see that $S_{n,\lambda}(f_\lambda) = S_{n,\lambda}(f_\lambda) - S_\lambda(f_\lambda) = \frac{1}{n} \sum_{i=1}^n R_i$. Let

$$\mathcal{E}_{n,1} = \{\|S_{n,\lambda}(f_\lambda)\|_f \leq L(M)r\},$$

then $\sup_{f \in H^m(C)} \mathbb{P}_f(\mathcal{E}_{n,1}^c) \leq 2 \exp(-Mnhr^2)$.

Define

$$\psi_{n,f}^{(1)}(X_i; g) = [C_2c_K]^{-1}h^{1/2}\ddot{A}(f_\lambda(X_i))g(X_i), \quad i = 1, \dots, n,$$

and $Z_{n,f}^{(1)}(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\psi_{n,f}^{(1)}(X_i; g)K_{X_i}^f - E_f\{\psi_{n,f}^{(1)}(X_i; g)K_{X_i}^f\}]$. It follows by Lemma S.5 that $\sup_{f \in H^m(C)} \mathbb{P}_f(\mathcal{E}_{n,2}^c) \leq 2 \exp(-Mnhr^2)$, where $\mathcal{E}_{n,2} = \{\sup_{g \in \mathcal{G}} \|Z_{n,f}^{(1)}(g)\|_f \leq \sqrt{Mnhr^2}B(h)\}$.

For any $g \in S^m(\mathbb{I}) \setminus \{0\}$, let $\bar{g} = g/d'_n$, where $d'_n = c_Kh^{-1/2}\|g\|_f$. It follows by Lemma S.1 that

$$\|\bar{g}\|_\infty \leq c_Kh^{-1/2}\|\bar{g}\|_f = c_Kh^{-1/2}\|g\|_f/d'_n = 1, \text{ and}$$

$$\begin{aligned} J(\bar{g}, \bar{g}) &= d_n'^{-2}J(g, g) \\ &= h^{-2m} \frac{\lambda J(g, g)}{c_K^2h^{-1}\|g\|_f^2} \leq h^{-2m} \frac{\|g\|^2}{c_K^2h^{-1}\|g\|_f^2} \leq 2C_2^2c_K^{-2}h^{-2m+1}. \end{aligned}$$

Therefore, $\bar{g} \in \mathcal{G}$. Consequently, on $\mathcal{E}_{n,2}$, for any $g \in S^m(\mathbb{I}) \setminus \{0\}$, we get $\|Z_{n,f}^{(1)}(\bar{g})\|_f \leq \sqrt{Mnhr^2}B(h)$, which leads to that

$$\begin{aligned} & \|DS_{n,\lambda}(f_\lambda)g - DS_\lambda(f_\lambda)g\|_f \\ &= \frac{1}{n} \left\| \sum_{i=1}^n [\ddot{A}(f_\lambda(X_i))g(X_i)K_{X_i}^f - E_f\{\ddot{A}(f_\lambda(X_i))g(X_i)K_{X_i}^f\}] \right\|_f \\ (S.12) \qquad \qquad \qquad & \leq C_2c_K^2M^{1/2}rh^{-1/2}B(h)\|g\|_f \leq \|g\|_f/6, \end{aligned}$$

where the last inequality follows by condition $C_2 c_K^2 M^{1/2} r h^{-1/2} B(h) \leq 1/6$. Note that the above inequality also holds for $g = 0$.

Next we define $T_{3f}(g) = S_{n,\lambda}(f_\lambda + g) - S_{n,\lambda}(f_\lambda) - DS_{n,\lambda}(f_\lambda)g$. Let $r_{2n} = 6L(M)r$. For any $g \in \mathcal{G}$ and $1 \leq i \leq n$, define $\tilde{\psi}_{n,i}(g) = |g(X_i)|$, and let $\tilde{Z}_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\tilde{\psi}_{n,i}(g) - E\{\tilde{\psi}_{n,i}(g)\}]$. It is easy to see that for any $g_1, g_2 \in \mathcal{G}$, $|\tilde{\psi}_{n,i}(g_1) - \tilde{\psi}_{n,i}(g_2)| \leq \|g_1 - g_2\|_\infty$. Following the proof of Lemma S.5 it can be shown that for any $t \geq 0$,

$$P\left(\sup_{g \in \mathcal{G}} |\tilde{Z}_n(g)| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{B(h)^2}\right),$$

and hence, we get that $P(\mathcal{E}_{n,3}^c) \leq 2 \exp(-Mnh r^2)$, where

$$\mathcal{E}_{n,3} = \{\sup_{g \in \mathcal{G}} |\tilde{Z}_n(g)| \leq \sqrt{Mnh r^2} B(h)\}.$$

On $\mathcal{E}_{n,2} \cap \mathcal{E}_{n,3}$, for any $g_1, g_2 \in \mathbb{B}(r_{2n})$ (with $g_1 \neq g_2$) and letting $g = g_1 - g_2$ (and hence $\|g_2 + sg\|_\infty \leq c_K h^{-1/2} r_{2n} \leq C/4$ for any $s \in [0, 1]$), together with $\|f_\lambda - f\|_\infty \leq C/4$, we have

$$\begin{aligned} & \|T_{3f}(g_1) - T_{3f}(g_2)\|_f \\ &= \|S_{n,\lambda}(f_\lambda + g_1) - S_{n,\lambda}(f_\lambda + g_2) - DS_{n,\lambda}(f_\lambda)g\|_f \\ &= \left\| \int_0^1 \int_0^1 D^2 S_{n,\lambda}(f_\lambda + s'(g_2 + sg))(g_2 + sg) g ds ds' \right\|_f \\ &\leq \int_0^1 \int_0^1 \|D^2 S_{n,\lambda}(f_\lambda + s'(g_2 + sg))(g_2 + sg)g\|_f ds ds' \\ &\leq \int_0^1 \int_0^1 \left\| \frac{1}{n} \sum_{i=1}^n \ddot{A}(f_\lambda(X_i) + s'(g_2(X_i) + sg(X_i))) \right. \\ &\quad \left. (g_2(X_i) + sg(X_i))g(X_i) K_{X_i}^f \right\|_f ds ds' \\ &\leq \int_0^1 \int_0^1 \frac{C_2}{n} \sum_{i=1}^n \|g_2 + sg\|_\infty \times |g(X_i)| \times \|K_{X_i}^f\|_f ds ds' \\ &\leq \frac{C_2 (c_K h^{-1/2})^2 r_{2n}}{n} \sum_{i=1}^n |g(X_i)| \\ &= \frac{C_2 (c_K h^{-1/2})^3 r_{2n}}{n} \left(\sum_{i=1}^n \tilde{\psi}_{n,i}(\bar{g}) \right) \|g\|_f, \end{aligned} \tag{S.13}$$

where $\bar{g} = g / (c_K h^{-1/2} \|g\|_f)$. Recalling the previous arguments we get $\bar{g} \in \mathcal{G}$. It can be shown by Cauchy-Schwartz inequality that

$$E\{\tilde{\psi}_{n,i}(\bar{g})\} = \frac{E\{|g(X_i)|\}}{c_K h^{-1/2} \|g\|_f} \leq \frac{C_2^{1/2} V_f(g, g)^{1/2}}{c_K h^{-1/2} \|g\|_f} \leq C_2^{1/2} c_K^{-1} h^{1/2}.$$

Since $\mathcal{E}_{n,3}$ implies $|\tilde{Z}_n(\bar{g})| \leq \sqrt{Mnh r^2} B(h)$, we get that

$$\frac{1}{n} \sum_{i=1}^n \tilde{\psi}_{n,i}(\bar{g}) \leq \sqrt{Mhr^2} B(h) + C_2^{1/2} c_K^{-1} h^{1/2}.$$

Therefore, (S.13) has upper bound

$$\begin{aligned}
 (\text{S.13}) \quad &\leq C_2(c_K h^{-1/2})^3 r_{2n}(\sqrt{Mhr^2}B(h) + C_2^{1/2}c_K^{-1}h^{1/2})\|g\|_f \\
 &= 12C_0C_2c_K^4(4C_1 + M)h^{-1}r(M^{1/2}rB(h) + C_2^{1/2}c_K^{-1})\|g\|_f \\
 (\text{S.14}) \quad &\leq \|g_1 - g_2\|_f/6,
 \end{aligned}$$

where the last inequality follows by condition

$$12C_0C_2c_K^4(4C_1 + M)h^{-1}r(M^{1/2}rB(h) + C_2^{1/2}c_K^{-1}) \leq 1/6.$$

Taking $g_2 = 0$ in (S.14) we get that $\|T_{3f}(g_1)\|_f \leq \|g_1\|_f/6$ for any $g_1 \in \mathbb{B}(r_{2n})$. Therefore, it follows by (S.12) that, for any $f \in H^m(C)$, on $\mathcal{E}_n := \mathcal{E}_{n,1} \cap \mathcal{E}_{n,2} \cap \mathcal{E}_{n,3}$ and for any $g \in \mathbb{B}(r_{2n})$,

$$\|T_{2f}(g)\|_f \leq 2(\|g\|_f/6 + \|g\|_f/6 + r_{2n}/6) \leq 2(r_{2n}/6 + r_{2n}/6 + r_{2n}/6) = r_{2n}.$$

Meanwhile, for any $g_1, g_2 \in \mathbb{B}(r_{2n})$, replacing g by $g_1 - g_2$ in (S.12), together with (S.13) and (S.14), we get that

$$\|T_{2f}(g_1) - T_{2f}(g_2)\|_f \leq 2(\|g_1 - g_2\|_f/6 + \|g_1 - g_2\|_f/6) = 2\|g_1 - g_2\|_f/3.$$

Therefore, for any $f \in H^m(C)$, on \mathcal{E}_n , T_{2f} is a contraction mapping from $\mathbb{B}(r_{2n})$ to itself. By contraction mapping theorem, there exists uniquely an element $g'' \in \mathbb{B}(r_{2n})$ s.t. $T_{2f}(g'') = g''$. Let $\hat{f}_{n,\lambda} = f_\lambda + g''$. Clearly, $S_{n,\lambda}(\hat{f}_{n,\lambda}) = 0$, and hence, $\hat{f}_{n,\lambda}$ is the maximizer of $\ell_{n,\lambda}$; see (5.2). So we get that, on \mathcal{E}_n , $\|\hat{f}_{n,\lambda} - f\|_f \leq \|f_\lambda - f\|_f + \|\hat{f}_{n,\lambda} - f_\lambda\|_f \leq r_{1n} + r_{2n} = 2bh^m + 6L(M)r$. The desired conclusion follows by the trivial fact: $\sup_{f \in H^m(C)} P_f(\mathcal{E}_n^c) \leq 6 \exp(-Mnhr^2)$. Proof of (a) is completed.

Next we show (b).

For any $f \in H^m(C)$, let $\hat{f}_{n,\lambda}$ be the penalized MLE of f obtained by (5.2). Let $g_n = \hat{f}_{n,\lambda} - f$, $\delta_n = 2bh^m + 6L(M)r$, $d'_n = c_K h^{-1/2} \delta_n$, and for $g \in \mathcal{G}$ define

$$\psi_{n,f}^{(2)}(X_i; g) = c_K^{-1} h^{1/2} [C_2 d'_n]^{-1} (\dot{A}(f(X_i) + d'_n g(X_i)) - \dot{A}(f(X_i))).$$

It can be seen that for any $g_1, g_2 \in \mathcal{G}$, by $\delta'_n = c_K h^{-1/2} \delta_n < C$, we have

$$\begin{aligned}
 &|\psi_{n,f}^{(2)}(X_i; g_1) - \psi_{n,f}^{(2)}(X_i; g_2)| \\
 &\leq c_K^{-1} h^{1/2} [C_2 d'_n]^{-1} C_2 d'_n \|g_1 - g_2\|_\infty = c_K^{-1} h^{1/2} \|g_1 - g_2\|_\infty.
 \end{aligned}$$

Let $\mathcal{E}_{n,4} = \{\sup_{g \in \mathcal{G}} \|Z_{n,f}^{(2)}(g)\|_f \leq \sqrt{Mnhr^2}B(h)\}$, where

$$Z_{n,f}^{(2)}(g) = \frac{1}{\sqrt{n}} \sum_{i \in I_j} [\psi_{n,f}^{(2)}(X_i; g) K_{X_i}^f - E_f^X \{\psi_{n,f}^{(2)}(X; g) K_X^f\}],$$

E_f^X denotes the expectation with respect to X (under P_f). It follows by Lemma S.5 that $\sup_{f \in H^m(C)} P_f(\mathcal{E}_{n,4}^c) \leq 2 \exp(-Mnhr^2)$.

On $\tilde{\mathcal{E}}_n := \mathcal{E}_n \cap \mathcal{E}_{n,4}$, we have $\|g_n\|_f \leq \delta_n$. Let $\bar{g} = g_n/d'_n$. Clearly, $\bar{g} \in \mathcal{G}$. Then we get that

$$\begin{aligned}
& \|S_{n,\lambda}(f + g_n) - S_{n,\lambda}(f) - (S_\lambda(f + g_n) - S_\lambda(f))\|_f \\
&= \frac{1}{n} \left\| \sum_{i=1}^n [(\dot{A}(f(X_i) + g_n(X_i)) - \dot{A}(f(X_i)))K_{X_i}^f \right. \\
&\quad \left. - E_f^X\{(\dot{A}(f(X) + g_n(X)) - \dot{A}(f(X)))K_X^f\}] \right\|_f \\
&= \frac{1}{n} \left\| \sum_{i \in I_j} [(\dot{A}(f(X_i) + d'_n \bar{g}(X_i)) - \dot{A}(f(X_i)))K_{X_i}^f \right. \\
&\quad \left. - E_f^X\{(\dot{A}(f(X) + d'_n \bar{g}(X)) - \dot{A}(f(X)))K_X^f\}] \right\|_f \\
&= \frac{C_2 c_K h^{-1/2} d'_n}{n} \left\| \sum_{i \in I_j} [\psi_{n,f}^{(2)}(X_i; \bar{g})K_{X_i}^f - E_f^X\{\psi_{n,f}^{(2)}(X; \bar{g})K_X^f\}] \right\|_f \\
&= \frac{C_2 c_K h^{-1/2} d'_n}{\sqrt{n}} \|Z_{n,f}^{(2)}(\bar{g})\|_f \leq C_2 c_K^2 M^{1/2} h^{-1/2} r B(h) \delta_n = a_n.
\end{aligned}
\tag{S.15}$$

It is easy to show that

$$\begin{aligned}
& \left\| \int_0^1 \int_0^1 s D^2 S_\lambda(f + ss'g_n) g_n g_n ds ds' \right\|_f \\
&= \left\| \int_0^1 \int_0^1 s E_f^X \{ \ddot{A}(f(X) + ss'g_n(X)) g_n(X)^2 K_X \} ds ds' \right\|_f \\
&\leq C_2 c_K h^{-1/2} \int_0^1 \int_0^1 s E_f^X \{ g_n(X)^2 \} ds ds' \\
&\leq C_2^2 c_K h^{-1/2} \|g_n\|_f^2 \leq C_2^2 c_K h^{-1/2} \delta_n^2 = b_n.
\end{aligned}
\tag{S.16}$$

Since $S_{n,\lambda}(f + g_n) = 0$ and $DS_\lambda(f) = -id$, from (S.15) and (S.16) we have on $\tilde{\mathcal{E}}_n$,

$$\begin{aligned}
a_n &\geq \|S_{n,\lambda}(f) + DS_\lambda(f)g_n + \int_0^1 \int_0^1 s D^2 S_\lambda(f + ss'g_n) g_n g_n ds ds'\|_f \\
&= \|S_{n,\lambda}(f) - g_n + \int_0^1 \int_0^1 s D^2 S_\lambda(f + ss'g_n) g_n g_n ds ds'\|_f \\
&\geq \|S_{n,\lambda}(f) - g_n\|_f - \left\| \int_0^1 \int_0^1 s D^2 S_\lambda(f + ss'g_n) g_n g_n ds ds' \right\|_f,
\end{aligned}$$

which implies that

$$\|\hat{f}_{n,\lambda} - f - S_{n,\lambda}(f)\|_f \leq a_n + b_n.$$

Since $\sup_{f \in H^m(C)} \mathbf{P}_f(\tilde{\mathcal{E}}_n^c) \leq 8 \exp(-Mnh r^2)$, proof of (b) is completed. \square

S.2. Proofs in Section 2.

PROOF OF LEMMA 3.1. Let $T : f \mapsto \{f_\nu : \nu \geq 1\}$ be the one-to-one map from $S^m(\mathbb{I})$ to \mathcal{R}_m , as defined in Lemma 5.2. Let $\tilde{\Pi}_\lambda$ and $\tilde{\Pi}$ be the probability measures induced by $\{w_\nu : \nu > m\}$

and $\{v_\nu : \nu > m\}$, respectively, which are both defined on \mathbb{R}^∞ . That is, for any subset $S \in \mathbb{R}^\infty$, $\tilde{\Pi}_\lambda(S) = \mathbb{P}(\{w_\nu : \nu > m\} \in S)$ and $\tilde{\Pi}(S) = \mathbb{P}(\{v_\nu : \nu > m\} \in S)$. Likewise, let Π'_λ and Π' be probability measures induced by $\{w_\nu : \nu \geq 1\}$ and $\{v_\nu : \nu \geq 1\}$. It is easy to see that, for any measurable $B \subseteq \mathcal{R}_m$,

$$\Pi_\lambda(T^{-1}B) = \mathbb{P}(G_\lambda \in T^{-1}B) = \mathbb{P}(\{w_\nu : \nu \geq 1\} \in B) = \Pi'_\lambda(B), \text{ and}$$

$$\Pi(T^{-1}B) = \mathbb{P}(G \in T^{-1}B) = \mathbb{P}(\{v_\nu : \nu \geq 1\} \in B) = \Pi'(B).$$

The following result can be found in Hájek [21].

Proposition S.2. *The Radon-Nikodym derivative of $\tilde{\Pi}_\lambda$ w.r.t. $\tilde{\Pi}$ is*

$$\begin{aligned} \frac{d\tilde{\Pi}_\lambda}{d\tilde{\Pi}}(\{f_\nu : \nu > m\}) &= \prod_{\nu>m}^{\infty} \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \exp\left(-\frac{n\lambda}{2}f_\nu^2\rho_\nu\right) \\ &= \prod_{\nu>m}^{\infty} \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \cdot \exp\left(-\frac{n\lambda}{2}\sum_{\nu>m}^{\infty} f_\nu^2\rho_\nu\right). \end{aligned}$$

Note that in Proposition S.2, $\prod_{\nu>m}^{\infty} \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2}$ is convergent since $\rho_\nu^{-\beta/(2m)} \asymp \nu^{-\beta}$. Therefore, by Proposition S.2, we have

$$\begin{aligned} &\frac{d\Pi'_\lambda}{d\Pi'}(\{f_\nu : \nu \geq 1\}) \\ &= \frac{\prod_{\nu=1}^m \left(\frac{2\pi\sigma_\nu^2}{1+n\lambda\sigma_\nu^2}\right)^{-1/2} \exp\left(-\frac{(1+n\lambda\sigma_\nu^2)f_\nu^2}{2\sigma_\nu^2}\right)}{\prod_{\nu=1}^m (2\pi\sigma_\nu^2)^{-1/2} \exp\left(-\frac{f_\nu^2}{2\sigma_\nu^2}\right)} \times \frac{d\tilde{\Pi}_\lambda}{d\tilde{\Pi}}(\{f_\nu : \nu > m\}) \\ &= \prod_{\nu=1}^m (1 + n\lambda\sigma_\nu^2)^{1/2} \exp\left(-\frac{n\lambda}{2}f_\nu^2\right) \times \prod_{\nu>m}^{\infty} \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \\ &\quad \times \exp\left(-\frac{n\lambda}{2}\sum_{\nu>m}^{\infty} f_\nu^2\rho_\nu\right) \\ &= \prod_{\nu=1}^m (1 + n\lambda\sigma_\nu^2)^{1/2} \prod_{\nu>m}^{\infty} \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \times \exp\left(-\frac{n\lambda}{2}\sum_{\nu=1}^{\infty} f_\nu^2\gamma_\nu\right). \end{aligned}$$

Then for any measurable $S \subseteq S^m(\mathbb{I})$, by change of variable, we have

$$\begin{aligned}
& \Pi_\lambda(S) \\
&= \Pi'_\lambda(TS) \\
&= \int_{TS} d\Pi'_\lambda(\{f_\nu : \nu \geq 1\}) \\
&= \prod_{\nu=1}^m (1 + n\lambda\sigma_\nu^2)^{1/2} \prod_{\nu>m}^\infty \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \\
&\quad \cdot \int_{TS} \exp\left(-\frac{n\lambda}{2} \sum_{\nu=1}^\infty f_\nu^2 \gamma_\nu\right) d\Pi'(\{f_\nu : \nu \geq 1\}) \\
&= \prod_{\nu=1}^m (1 + n\lambda\sigma_\nu^2)^{1/2} \prod_{\nu>m}^\infty \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \\
&\quad \cdot \int_{TS} \exp\left(-\frac{n\lambda}{2} J(T^{-1}(\{f_\nu : \nu \geq 1\}))\right) d(\Pi \circ T^{-1})(\{f_\nu : \nu \geq 1\}) \\
&= \prod_{\nu=1}^m (1 + n\lambda\sigma_\nu^2)^{1/2} \prod_{\nu>m}^\infty \left(1 + n\lambda\rho_\nu^{-\beta/(2m)}\right)^{1/2} \int_S \exp\left(-\frac{n\lambda}{2} J(f)\right) d\Pi(f).
\end{aligned}$$

This completes the proof of the lemma. \square

S.3. *Proofs in Section 4.* Proof of Theorem 4.1 requires the following result.

Theorem A.1. (An initial contraction rate) Under Assumption A1, if $r_n = o(h^{3/2})$, $h^{1/2} \log N = o(1)$, $nh^{2m+1} \geq 1$, and $f_0 = \sum_{\nu=1}^\infty f_\nu^0 \varphi_\nu$ satisfies Condition (S), then there exists a universal constant $M > 0$ s.t. $P(\|f - f_0\| \geq Mr_n | \mathbf{D}_j) = o_{P_{f_0}^n}(1)$ as $n \rightarrow \infty$.

Before proving Theorem A.1, we present a preliminary lemma.

Let $\{\tilde{\varphi}_\nu : \nu \geq 1\}$ be a bounded orthonormal basis of $L^2(\mathbb{I})$ under usual L^2 inner product. For any $b \in [0, \beta]$, define

$$\tilde{H}_b = \left\{ \sum_{\nu=1}^\infty f_\nu \tilde{\varphi}_\nu : \sum_{\nu=1}^\infty f_\nu^2 \rho_\nu^{1+b/(2m)} < \infty \right\}.$$

Then \tilde{H}_b can be viewed as a version of Sobolev space with regularity $m + b/2$. Define $\tilde{G} = \sum_{\nu=1}^\infty v_\nu \tilde{\varphi}_\nu$, a centered GP, and $\tilde{f}_0 = \sum_{\nu=1}^\infty f_\nu^0 \tilde{\varphi}_\nu$. Define $\tilde{V}(f, g) = \langle f, g \rangle_{L^2} = \int_0^1 f(x)g(x)dx$, the usual L^2 inner product, $\tilde{J}(f) = \sum_{\nu=1}^\infty |\tilde{V}(f, \tilde{\varphi}_\nu)|^2 \rho_\nu$, a functional on \tilde{H}_0 . For simplicity, denote $\tilde{V}(f) = \tilde{V}(f, f)$. Clearly, $\tilde{f}_0 \in \tilde{H}_\beta$. Since \tilde{G} is a Gaussian process with covariance function

$$\tilde{R}(s, t) = E\{\tilde{G}(s)\tilde{G}(t)\} = \sum_{\nu=1}^m \sigma_\nu^2 \tilde{\varphi}_\nu(s) \tilde{\varphi}_\nu(t) + \sum_{\nu>m} \rho_\nu^{-(1+\frac{\beta}{2m})} \tilde{\varphi}_\nu(s) \tilde{\varphi}_\nu(t),$$

it follows by [52] that \tilde{H}_β is the RKHS of \tilde{G} . For any \tilde{H}_b with $0 \leq b \leq \beta$, define inner product

$$\left\langle \sum_{\nu=1}^\infty f_\nu \tilde{\varphi}_\nu, \sum_{\nu=1}^\infty g_\nu \tilde{\varphi}_\nu \right\rangle_b = \sum_{\nu=1}^m \sigma_\nu^{-2} f_\nu g_\nu + \sum_{\nu>m} f_\nu g_\nu \rho_\nu^{1+\frac{b}{2m}}.$$

Let $\|\cdot\|_b$ be the norm corresponding to the above inner product. The following lemma is used in the proof of Theorem A.1.

Lemma S.7. *Let d_n be any positive sequence. If Condition (S) holds, then there exists $\omega \in \tilde{H}_\beta$ such that*

- (i). $\tilde{V}(\omega - \tilde{f}_0) \leq \frac{1}{4}d_n^2$,
- (ii). $\tilde{J}(\omega - \tilde{f}_0) \leq \frac{1}{4}d_n^{\frac{2(\beta-1)}{2m+\beta-1}}$,
- (iii). $\|\omega\|_\beta^2 = O(d_n^{\frac{2}{2m+\beta-1}})$.

PROOF OF LEMMA S.7. Let $\omega = \sum_{\nu=1}^{\infty} \omega_\nu \tilde{\varphi}_\nu$, where $\omega_\nu = \frac{df_\nu^0}{d+(\sigma b_\nu)^\alpha}$, $\sigma = d_n^{2/(2m+\beta-1)}$, $b_\nu = \rho_\nu^{1/(2m)}$, $\alpha = m + (\beta - 1)/2$, and $d > 0$ is a constant to be described. It is easy to see that for any ν , $f_\nu^0 - \omega_\nu = \frac{(\sigma b_\nu)^\alpha f_\nu^0}{d+(\sigma b_\nu)^\alpha}$. Then

$$\begin{aligned} \tilde{V}(\omega - \tilde{f}_0) &= \sum_{\nu=1}^{\infty} (f_\nu^0 - \omega_\nu)^2 = \sum_{\nu=1}^{\infty} \frac{|f_\nu^0|^2 (\sigma b_\nu)^{2\alpha}}{(d + (\sigma b_\nu)^\alpha)^2} \\ &\leq \sigma^{2m+\beta-1} d^{-2} \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \rho_\nu^{1+\frac{\beta-1}{2m}}, \end{aligned}$$

and

$$\begin{aligned} \tilde{J}(\omega - \tilde{f}_0) &= \sum_{\nu=1}^{\infty} (f_\nu^0 - \omega_\nu)^2 \rho_\nu \\ &= \sigma^{\beta-1} \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \rho_\nu^{1+\frac{\beta-1}{2m}} (d(\sigma b_\nu)^{-m} + (\sigma b_\nu)^{(\beta-1)/2})^{-2} \\ &\leq d^{-\frac{\beta-1}{k}} \left(\left(\frac{2m}{\beta-1} \right)^{-\frac{m}{k}} + \left(\frac{2m}{\beta-1} \right)^{\frac{\beta-1}{2k}} \right)^{-2} \sigma^{\beta-1} \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \rho_\nu^{1+\frac{\beta-1}{2m}}, \end{aligned}$$

where in the last equation $k = m + (\beta - 1)/2$. Therefore, we choose d as a suitably large fixed constant such that (i) and (ii) hold.

To show (iii), observe that

$$\begin{aligned} \|\omega\|_\beta^2 &= \sum_{\nu=1}^m \sigma_\nu^{-2} \omega_\nu^2 + \sum_{\nu>m} \omega_\nu^2 \rho_\nu^{1+\frac{\beta}{2m}} = \sum_{\nu=1}^m \sigma_\nu^{-2} |f_\nu^0|^2 + \sum_{\nu>m} \frac{d^2 |f_\nu^0|^2 \rho_\nu^{1+\frac{\beta-1}{2m}}}{(d + (\sigma b_\nu)^\alpha)^2} b_\nu \\ &= O(\sigma^{-1}). \end{aligned}$$

The result follows by $\sigma = d_n^{2/(2m+\beta-1)}$. □

PROOF OF THEOREM A.1. Note that there exists a universal constant $c' > 0$ such that $\Psi(x) \leq c' x^{1-1/(2m)}$ for any $0 < x < 1$. Therefore, there exists a universal constant $c'' > 0$ s.t. $B(h) \leq c'' h^{-(2m-1)/(4m)}$.

To prove the theorem, we first show the following posterior consistency: for any $\varepsilon > 0$, as $n \rightarrow \infty$,

$$(S.17) \quad P(\|f - f_0\|_\infty \geq \varepsilon | \mathbf{D}_n) \rightarrow 0, \text{ in } \mathbb{P}_{f_0}^n\text{-probability.}$$

We can rewrite the posterior density of f by

$$p(f | \mathbf{D}_n) = \frac{\prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-n\lambda J(f)/2) d\Pi(f)}{\int_{S^m(\mathbb{I})} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-n\lambda J(f)/2) d\Pi(f)},$$

where recall that $p_f(z)$ is the probability density of $Z = (Y, X)$ under f .

First of all, we give a lower bound for

$$I_1 = \int_{S^m(\mathbb{I})} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-n\lambda J(f)/2) d\Pi(f).$$

Define $B_n = \{f \in S^m(\mathbb{I}) : V(f - f_0) \leq r_n^2, J(f - f_0) \leq r_n^{\frac{2(\beta-1)}{2m+\beta-1}}\}$. Then

$$\begin{aligned} I_1 &\geq \int_{B_n} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-n\lambda J(f)/2) d\Pi(f) \\ &= \int_{B_n} \exp\left(\sum_{i=1}^n R_i(f, f_0)\right) \exp(-n\lambda J(f)/2) d\Pi(f), \end{aligned}$$

where $R_i(f, f_0) = \log(p_f(Z_i)/p_{f_0}(Z_i)) = Y_i(f(X_i) - f_0(X_i)) - A(f(X_i)) + A(f_0(X_i))$ for any $1 \leq i \leq n$. Define $d\Pi^*(f) = d\Pi(f)/\Pi(B_n)$, a reduced probability measure on B_n . By Jensen's inequality,

$$\begin{aligned} &\log \int_{B_n} \exp\left(\sum_{i=1}^n R_i(f, f_0)\right) \exp(-n\lambda J(f)/2) d\Pi^*(f) \\ &\geq \int_{B_n} \left(\sum_{i=1}^n R_i(f, f_0) - n\lambda J(f)/2\right) d\Pi^*(f) \\ &= \int_{B_n} \sum_{i=1}^n [R_i(f, f_0) - E_{f_0}\{R_i(f, f_0)\}] d\Pi^*(f) \\ &\quad + n \int_{B_n} E_{f_0}\{R_i(f, f_0)\} d\Pi^*(f) - \int_{B_n} \frac{n\lambda J(f)}{2} d\Pi^*(f) \\ &\equiv J_1 + J_2 + J_3. \end{aligned}$$

For any $f \in B_n$, $\|f - f_0\|^2 = V(f - f_0) + \lambda J(f - f_0) \leq r_n^2 + \lambda r_n^{\frac{2(\beta-1)}{2m+\beta-1}}$. By Lemma S.4 and the condition $h^{-3/2}r_n = o(1)$, we can choose n to be sufficiently large so that $\|f - f_0\|_\infty \leq ch^{-1/2}\|f - f_0\| \leq c\sqrt{h^{-1}r_n^2 + h^{2m-1}} \leq 1$.

It follows from Assumption A1 that for $C = 1 + C_3\sqrt{J(f_0)}$, there exist positives C'_0, C'_1, C'_2 s.t. (2.2) and (2.3) hold with C_0, C_1, C_2 therein replaced by C'_0, C'_1, C'_2 , respectively.

It follows by Taylor's expansion, $E_{f_0}\{Y_i - \dot{A}(f_0(X_i))|X_i\} = 0$, $\ddot{A}(z) \leq C'_2$ for $|z| \leq 2C$ and Assumption A1 that for any $f \in B_n$,

$$|E_{f_0}\{R_i(f, f_0)\}| \leq C'_2 E_{f_0}\{(f(X) - f_0(X))^2\} \leq (C'_2)^2 r_n^2.$$

Therefore, $J_2 \geq -(C'_2)^2 n r_n^2$.

Since $r_n^2 = o(1)$, we can choose n to be large so that $|E_{f_0}\{R_i(f, f_0)\}| \leq 1$. Meanwhile, for any $f \in B_n$, for some $s \in [0, 1]$, we have

$$\begin{aligned} & |R_i(f, f_0)| \\ &= |Y_i(f(X_i) - f_0(X_i)) - A(f(X_i)) + A(f_0(X_i))| \\ &= |Y_i - \dot{A}(f_0(X_i)) \\ &\quad - \frac{1}{2}\ddot{A}(f_0(X_i) + s(f(X_i) - f_0(X_i)))(f - f_0)(X_i)| \times |(f - f_0)(X_i)| \\ &\leq |Y_i - \dot{A}(f_0(X_i))| + C'_2/2. \end{aligned}$$

We have used $\|f - f_0\|_\infty \leq 1$ in the above inequalities.

For any $1 \leq i \leq n$, define $A_i = \{|Y_i - \dot{A}(f_0(X_i))| \leq 2C'_0 \log n\}$. It follows by Assumption A1 that $\mathbf{P}_{f_0}^n(\cup_{i=1}^n A_i^c) \leq C'_1/n \rightarrow 0$, as $n \rightarrow \infty$. Define $\xi_i = \int_{B_n} R_i(f, f_0) d\Pi^*(f) \times I_{A_i}$, we get that $|\xi_i| \leq 2C'_0 \log n + C'_2/2$, a.s. It can also be shown by $r_n^2 \geq 1/n$ that

$$\begin{aligned} & |E_{f_0}\{\int_{B_n} R_i(f, f_0) d\Pi^*(f) \times I_{A_i^c}\}| \\ &\leq E_{f_0}\{(|Y_i - f_0(X_i)| + C'_2/2) \times I_{A_i^c}\} \\ &= E_{f_0}\{|Y_i - f_0(X_i)| \times I_{A_i^c}\} + \frac{C'_2}{2} \mathbf{P}_{f_0}^n(A_i^c) \\ &\leq C'_0 \sqrt{2C'_1 \mathbf{P}_{f_0}^n(A_i^c)^{1/2}} + \frac{C'_2}{2} \mathbf{P}_{f_0}^n(A_i^c) \\ &\leq \frac{\sqrt{2}C'_0 C'_1}{n} + \frac{C'_1 C'_2}{2n^2} \leq (\sqrt{2}C'_0 C'_1 + C'_1 C'_2) r_n^2. \end{aligned}$$

Let $\delta = 1/(\sqrt{n}r_n)$. Note that by the condition $h^{1/2} \log n = o(1)$ we have $\delta \log n = (\log n)/(\sqrt{n}r_n) \leq h^{1/2} \log n = o(1)$, we can let n be large so that $\delta(4C'_0 \log n + C'_2) \leq 1$. Let $d_i = \xi_i - E_{f_0}\{\xi_i\}$ for $1 \leq i \leq n$, then it is easy to see that

$$|d_i| \leq |\xi_i| + |E_{f_0}\{\xi_i\}| \leq 4C'_0 \log n + C'_2, \text{ a.s.}$$

Let $e_i = E_{f_0}\{\exp(\delta|d_i|) - 1 - \delta|d_i|\}$. It can be shown using inequality $\exp(x) - 1 - x \leq x^2 \exp(x)$

for $x \geq 0$ and Cauchy-Schwartz inequality that

$$\begin{aligned}
|e_i| &\leq E_{f_0} \{ \delta^2 d_i^2 \exp(\delta |d_i|) \} \\
&\leq e \delta^2 E_{f_0} \{ d_i^2 \} \\
&\leq e \delta^2 E_{f_0} \{ \xi_i^2 \} \\
&\leq e \delta^2 \int_{B_n} E_{f_0} \{ R_i(f, f_0)^2 \} d\Pi^*(f) \\
&\leq e \delta^2 \int_{B_n} E_{f_0} \{ (|Y_i - \dot{A}(f_0(X_i))| + C'_2/2)^2 (f - f_0)(X_i)^2 \} d\Pi^*(f) \\
&\leq e(4(C'_0)^2 C'_1 C'_2 + (C'_2)^3) \delta^2 r_n^2,
\end{aligned}$$

where the last step follows from $V(f - f_0) \leq r_n^2$ for any $f \in B_n$. Therefore, it follows by [37, Theorem 3.2] that

$$\begin{aligned}
&\mathbf{P}_{f_0}^n \left(\left| \sum_{i=1}^n [\xi_i - E_{f_0} \{ \xi_i \}] \right| \geq (e(4(C'_0)^2 C'_1 C'_2 + (C'_2)^3) + 2) \sqrt{n} r_n \log n \right) \\
&\leq 2 \exp(- (e(4(C'_0)^2 C'_1 C'_2 + (C'_2)^3) + 2) \sqrt{n} r_n (\log n) \delta \\
&\quad + e(4(C'_0)^2 C'_1 C'_2 + (C'_2)^3) \delta^2 n r_n^2) \\
&\leq 2/n^2 \rightarrow 0, \text{ as } n \rightarrow \infty.
\end{aligned}$$

(S.18)

Since $\sqrt{n} r_n \gg \log n$, we can let n be large so that $(e(4(C'_0)^2 C'_1 C'_2 + (C'_2)^3) + 2) \sqrt{n} r_n \log n \leq n r_n^2$. Since on $\cap_{i=1}^n A_i$,

$$J_1 = \sum_{i=1}^n [\xi_i - E_{f_0} \{ \xi_i \}] - n E_{f_0} \left\{ \int_{B_n} R_i(f, f_0) d\Pi^*(f) \times I_{A_i^c} \right\},$$

we get from (S.18) that with $\mathbf{P}_{f_0}^n$ -probability approaching one,

$$J_1 \geq - (e(4(C'_0)^2 C'_1 C'_2 + (C'_2)^3) + 2) \sqrt{n} r_n \log n - n r_n^2 \geq -2n r_n^2.$$

Meanwhile, for any $f \in B_n$, $J(f) \leq (1 + J(f_0)^{1/2})^2$. Therefore, $J_3 \geq -\frac{(1+J(f_0)^{1/2})^2}{2} n \lambda$. So, with probability approaching one,

$$I_1 \geq \exp \left(- (2 + (C'_2)^2) n r_n^2 - \frac{(1 + J(f_0)^{1/2})^2}{2} n \lambda \right) \Pi(B_n).$$

To proceed, we need a lower bound for $\Pi(B_n)$. It follows by Lemma S.7 by replacing d_n therein by r_n , by Gaussian correlation inequality (see Theorem 1.1 of [32]), by Cameron-Martin theorem

(see [6] or [30, eqn (4.18)]) and [23, Example 4.5] that

$$\begin{aligned}
\Pi(B_n) &= P(V(G - f_0) \leq r_n^2, J(G - f_0) \leq r_n^{\frac{2(\beta-1)}{2m+\beta-1}}) \\
&= P(\tilde{V}(\tilde{G} - \tilde{f}_0) \leq r_n^2, \tilde{J}(\tilde{G} - \tilde{f}_0) \leq r_n^{\frac{2(\beta-1)}{2m+\beta-1}}) \\
&\geq P(\tilde{V}(\tilde{G} - \omega) \leq r_n^2/4, \tilde{J}(\tilde{G} - \omega) \leq r_n^{\frac{2(\beta-1)}{2m+\beta-1}}/4) \\
&\geq \exp(-\frac{1}{2}\|\omega\|_\beta^2) P(\tilde{V}(\tilde{G}) \leq r_n^2/4, \tilde{J}(\tilde{G}) \leq r_n^{\frac{2(\beta-1)}{2m+\beta-1}}/4) \\
&\geq \exp(-\frac{1}{2}\|\omega\|_\beta^2) P(\tilde{V}(\tilde{G}) \leq r_n^2/8) P(\tilde{J}(\tilde{G}) \leq r_n^{\frac{2(\beta-1)}{2m+\beta-1}}/8) \\
&\geq \exp(-c_1 r_n^{-2/(2m+\beta-1)}),
\end{aligned}
\tag{S.19}$$

where $c_1 > 0$ is a universal constant.

Since $\beta > 1$ and $r_n^2 = (nh)^{-1} + \lambda \geq n^{-2m/(2m+1)}$, we get $r_n^2 \geq \lambda$ and $nr_n^{\frac{2(2m+\beta)}{2m+\beta-1}} \geq n^{1 - \frac{2m(2m+\beta)}{(2m+1)(2m+\beta-1)}} > 1$, so $nr_n^2 > r_n^{-\frac{2}{2m+\beta-1}}$. Consequently, with $\mathbf{P}_{f_0}^n$ -probability approaching one,

$$I_1 \geq \exp(-c_2 nr_n^2), \tag{S.20}$$

where $c_2 = 2 + (C'_2)^2 + (1 + J(f_0)^{1/2})^2/2 + c_1$.

Now we choose a different constant C :

$$C = \max\{2C_3\sqrt{c_2 + 1}, c_2 + 1, 2(1 + C_3\sqrt{J(f_0)})\}. \tag{S.21}$$

It follows by Assumption A1 that there exist positives C_0, C_1, C_2 s.t. (2.2) and (2.3) hold. Next we examine

$$I_2 := \int_{A_n} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-\frac{n\lambda}{2} J(f)) d\Pi(f),$$

where $A_n = \{f \in S^m(\mathbb{I}) : \|f - f_0\| \geq 3C_2\delta_n\}$, $\delta_n = 2bh^m + 24C_0c_K(C)(4C_1 + C)r$, $r = r_n h^{-1/2}$, and $b = \frac{C_2C}{C_3} \sqrt{1 + \frac{1}{\rho_{m+1}}}$. By the condition $h^{-3/2}r_n = o(1)$ and $B(h) \lesssim h^{-(2m-1)/(4m)}$ it can be easily checked that the Rate Condition (H): (i)–(iv) are satisfied (when n becomes large) with M therein replaced by C . Define test $\phi_n = I(\|\hat{f}_{n,\lambda} - f_0\| \geq C_2\delta_n)$. Since $C_2 \geq 1$, it follows by part (a) of Theorem S.6 that

$$E_{f_0}\{\phi_n\} = \mathbf{P}_{f_0}^n(\|\hat{f}_{n,\lambda} - f_0\| \geq C_2\delta_n) \leq \mathbf{P}_{f_0}^n(\|\hat{f}_{n,\lambda} - f_0\| \geq \delta_n) \leq 6 \exp(-Cnr_n^2),$$

and by (S.4),

$$\begin{aligned}
\sup_{\substack{f \in H^m(C) \\ \|f - f_0\| \geq 3C_2\delta_n}} E_f\{1 - \phi_n\} &= \sup_{\substack{f \in H^m(C) \\ \|f - f_0\| \geq 3C_2\delta_n}} \mathbf{P}_f^n(\|\hat{f}_{n,\lambda} - f_0\| < C_2\delta_n) \\
&\leq \sup_{\substack{f \in H^m(C) \\ \|f - f_0\| \geq 3C_2\delta_n}} \mathbf{P}_f^n(\|\hat{f}_{n,\lambda} - f\| \geq 2C_2\delta_n) \\
&\leq \sup_{\substack{f \in H^m(C) \\ \|f - f_0\| \geq 3C_2\delta_n}} \mathbf{P}_f^n(\|\hat{f}_{n,\lambda} - f\|_f \geq \delta_n) \\
&\leq 6 \exp(-Cnr_n^2),
\end{aligned}$$

where the second last inequality follows by Lemma S.1.

Note that for any $f \in A_n \setminus H^m(C)$,

$$J(f) > (1 + 1/\rho_{m+1})^{-1} C_2^{-2} b^2 = C^2/C_3^2 \geq 4(c_2 + 1).$$

Since $nh^{2m+1} \geq 1$ leads to $r_n^2 = (nh)^{-1} + \lambda \leq 2\lambda$, it then holds that,

$$\begin{aligned} & E_{f_0} \{I_2(1 - \phi_n)\} \\ &= \int_{A_n} E_f \{1 - \phi_n\} \exp(-n\lambda J(f)/2) d\Pi(f) \\ &= \int_{A_n \setminus H^m(C)} E_f \{1 - \phi_n\} \exp(-n\lambda J(f)/2) d\Pi(f) \\ &\quad + \int_{A_n \cap H^m(C)} E_f \{1 - \phi_n\} \exp(-n\lambda J(f)/2) d\Pi(f) \\ &\leq \exp(-2n\lambda(c_2 + 1)) + 6 \exp(-(c_2 + 1)nr_n^2) \\ &\leq \exp(-(c_2 + 1)nr_n^2) + 6 \exp(-(c_2 + 1)nr_n^2) = 7 \exp(-(c_2 + 1)nr_n^2), \end{aligned}$$

so

$$E_{f_0} \{I_2(1 - \phi_n)\} \leq 7 \exp(-(c_2 + 1)nr_n^2),$$

which implies $I_2(1 - \phi_n) = O_{\mathbf{P}_{f_0}^n}(\exp(-(c_2 + 1)nr_n^2))$. On the other hand,

$$E_{f_0} \{P(A_n | \mathbf{D}_n) \phi_n\} \leq \mathbf{P}_{f_0}^n(\|\hat{f}_{n,\lambda} - f_0\| \geq C_2 \delta_n) \leq 6 \exp(-(c_2 + 1)nr_n^2),$$

so as $n \rightarrow \infty$,

$$E_{f_0} \{P(A_n | \mathbf{D}_n) \phi_n\} \leq 6 \exp(-(c_2 + 1)nr_n^2) \rightarrow 0,$$

which implies that $P(A_n | \mathbf{D}_n) \phi_n = o_{\mathbf{P}_{f_0}^n}(1)$. By the above arguments and (S.20), we have

$$\begin{aligned} & P(A_n | \mathbf{D}_n) \\ &= P(A_n | \mathbf{D}_n) \phi_n + P(A_n | \mathbf{D}_n) (1 - \phi_n) \\ &\leq P(A_n | \mathbf{D}_n) \phi_n + \frac{I_2(1 - \phi_n)}{I_1} \\ &= o_{\mathbf{P}_{f_0}^n}(1) + O_{\mathbf{P}_{f_0}^n}(\exp(-(c_2 + 1)nr_n^2) \exp(c_2 nr_n^2)) = o_{\mathbf{P}_{f_0}^n}(1), \end{aligned}$$

where the last step follows from $\exp(-nr_n^2) \leq \exp(-h^{-1}) = o(1)$. By condition $r_n h^{-3/2} = o(1)$ and the trivial fact $\delta_n \asymp r_n h^{-1/2}$, we have that $h^{-1/2} \delta_n = o(1)$, together with Lemma S.4 we have that (S.17) holds.

To prove the theorem, we let

$$I'_2 := \int_{A'_n} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-\frac{n\lambda}{2} J(f)) d\Pi(f),$$

where $A'_n = \{f \in S^m(\mathbb{I}) : \|f - f_0\| \geq \sqrt{2}Mr_n\}$ for a fixed number

$$M > \max\{2, J(f_0)^{1/2} + \sqrt{2(c_2 + 1)}, 1 + \|f_0\|_\infty\}$$

to be further described later. Let

$$A'_{n1} = \{f \in S^m(\mathbb{I}) : V(f - f_0) \geq M^2r_n^2, \lambda J(f - f_0) \leq M^2r_n^2\}$$

and

$$A'_{n2} = \{f \in S^m(\mathbb{I}) : \lambda J(f - f_0) \geq M^2r_n^2\}.$$

For any $f \in A'_{n2}$, it can be shown that

$$Mr_n \leq \sqrt{\lambda J(f - f_0)} \leq \sqrt{\lambda}(J(f)^{1/2} + J(f_0)^{1/2}) \leq (\lambda J(f))^{1/2} + J(f_0)^{1/2}r_n,$$

which leads to $\lambda J(f) \geq (M - J(f_0)^{1/2})^2r_n^2$. So we have

$$\begin{aligned} & E_{f_0} \left\{ \int_{A'_{n2}} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp\left(-\frac{n\lambda}{2}J(f)\right) d\Pi(f) \right\} \\ &= \int_{A'_{n2}} \exp\left(-\frac{n\lambda}{2}J(f)\right) d\Pi(f) \leq \exp(-(M - J(f_0)^{1/2})^2nr_n^2/2), \end{aligned}$$

which leads to that

$$\begin{aligned} & \int_{A'_{n2}} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp\left(-\frac{n\lambda}{2}J(f)\right) d\Pi(f) \\ (S.22) \quad &= O_{P_{f_0}^n}(\exp(-(M - J(f_0)^{1/2})^2nr_n^2/2)). \end{aligned}$$

To continue, we need to build uniformly consistent test. Let $d_H^2(P_f, P_g) = \frac{1}{2} \int (\sqrt{dP_f} - \sqrt{dP_g})^2$ be the squared Hellinger distance between the two probability measures $P_f(z)$ and $P_g(z)$. Recall that their corresponding probability density functions are p_f and p_g , respectively. Next we present a lemma showing the local equivalence of V and d_H^2 .

Lemma S.8. *Let C be chosen as (S.21) and C_0, C_1, C_2 be positives satisfying Assumption A1. Let $\varepsilon > 0$ satisfy $\varepsilon < \min\{1, 1/C_0, C\}$ and*

$$\frac{1}{12}C_2^2\varepsilon + \frac{1}{32}C_2^3\varepsilon^2 + C_0^3C_1C_2\varepsilon \exp\left(\frac{\varepsilon}{4}C_2 + \frac{C_2}{4C_0}\right) < \frac{1}{16}.$$

Then for any $f, g \in \mathcal{F}(C)$ satisfying $\|f - g\|_\infty \leq \varepsilon$,

$$V(f - g)/16 \leq d_H^2(P_f, P_g) \leq 3V(f - g)/16.$$

PROOF OF LEMMA S.8. For any $f, g \in \mathcal{F}(C)$ with $\|f - g\|_\infty \leq \varepsilon$, define $\Delta_Z(f, g) = \frac{1}{2}[Y(f(X) - g(X)) - A(f(X)) + A(g(X))]$, where recall and $Z = (Y, X)$. It is easy to see by direct calculations that

$$d_H^2(P_f, P_g) = 1 - E_g\{\exp(\Delta_Z(f, g))\}.$$

By Taylor's expansion, for some random $t \in [0, 1]$,

$$\begin{aligned} & 1 - E_g\{\exp(\Delta_Z(f, g))\} \\ &= -E_g\{\Delta_Z(f, g)\} - \frac{1}{2}E_g\{\Delta_Z(f, g)^2\} - \frac{1}{6}E_g\{\exp(t\Delta_Z(f, g))\Delta_Z(f, g)^3\}. \end{aligned}$$

We will analyze the terms on the right side of the equation.

Define $\xi = Y - \dot{A}(g(X))$. By [33] we get $E_g\{\xi|X\} = 0$ and $E_g\{\xi^2|X\} = \ddot{A}(g(X))$. By Taylor's expansion,

$$\begin{aligned} \Delta_Z(f, g) &= \frac{1}{2}[\xi(f(X) - g(X)) - \frac{1}{2}\ddot{A}(g(X))(f(X) - g(X))^2 \\ &\quad - \frac{1}{6}\ddot{A}(f_{1*}(X))(f(X) - g(X))^3], \end{aligned}$$

$$\Delta_Z(f, g) = \frac{1}{2}[\xi(f(X) - g(X)) - \frac{1}{2}\ddot{A}(f_{2*}(X))(f(X) - g(X))^2],$$

where $f_{k*}(X)$ is between $g(X)$ and $f(X)$ for $k = 1, 2$. It clearly holds that $\|f_{k*}\|_\infty \leq \|f\|_\infty + \|g - f\|_\infty < 2C$. Then we get that

$$-E_g\{\Delta_Z(f, g)\} = \frac{1}{4}V(f - g) + \frac{1}{12}E_g\{\ddot{A}(f_{1*}(X))(f(X) - g(X))^3\},$$

and

$$\begin{aligned} & E_g\{\Delta_Z(f, g)^2\} \\ &= E_g\left\{\left(\frac{1}{2}\xi(f(X) - g(X)) - \frac{1}{4}\ddot{A}(f_{2*}(X))(f(X) - g(X))^2\right)^2\right\} \\ &= \frac{1}{4}E_g\{\xi^2(f(X) - g(X))^2\} - \frac{1}{4}E_g\{\xi(f(X) - g(X))^3\ddot{A}(f_{2*}(X))\} \\ &\quad + \frac{1}{16}E_g\{\ddot{A}(f_{2*}(X))^2(f(X) - g(X))^4\} \\ &= \frac{1}{4}V(f - g) + \frac{1}{16}E_g\{\ddot{A}(f_{2*}(X))^2(f(X) - g(X))^4\}. \end{aligned}$$

Since $\|f - g\|_\infty \leq \varepsilon < \min\{1, 1/C_0, C\}$ and $0 < \ddot{A}(z) \leq C_2$ for any $z \in [-2C, 2C]$, implying $|\Delta_Z(f, g)| \leq \frac{1}{2}(|\xi| + C_2/2)|f(X) - g(X)|$, we get

$$\begin{aligned} & |E_g\{\exp(t\Delta_Z(f, g))\Delta_Z(f, g)^3\}| \\ &\leq E_g\{\exp(|\Delta_Z(f, g)|)|\Delta_Z(f, g)|^3\} \\ &\leq E_g\{\exp(\varepsilon|\xi|/2 + C_2\varepsilon/4)(|\xi|/2 + C_2/4)^3|f(X) - g(X)|^3\} \\ &= 6C_0^3E_g\left\{\exp(\varepsilon|\xi|/2 + C_2\varepsilon/4) \times \frac{1}{3!}\left(\frac{|\xi|/2 + C_2/4}{C_0}\right)^3|f(X) - g(X)|^3\right\} \\ &\leq 6C_0^3E_g\{\exp(\varepsilon|\xi|/2 + C_2\varepsilon/4)\exp(|\xi|/(2C_0) + C_2/(4C_0))|f(X) - g(X)|^3\} \\ &\leq 6C_0^3\exp(C_2\varepsilon/4 + C_2/(4C_0))E_g\{\exp(|\xi|/C_0)|f(X) - g(X)|^3\} \\ &\leq 6C_0^3C_1C_2\exp(C_2\varepsilon/4 + C_2/(4C_0))\varepsilon V(f - g). \end{aligned}$$

It also holds that

$$\begin{aligned} |E_g\{\ddot{A}(f_{1*}(X))(f(X) - g(X))^3\}| &\leq C_2^2 \varepsilon V(f - g), \\ |E_g\{\ddot{A}(f_{2*}(X))^2(f(X) - g(X))^4\}| &\leq C_2^3 \varepsilon^2 V(f - g). \end{aligned}$$

Therefore, by the above argument it holds that, for any $f, g \in \mathcal{F}(C)$ with $\|f - g\|_\infty \leq \varepsilon$,

$$\begin{aligned} &|d_H^2(\mathbf{P}_f, \mathbf{P}_g) - V(f - g)/8| \\ = & \left| \frac{1}{12} E_g\{\ddot{A}(f_{1*}(X))(f(X) - g(X))^3\} \right. \\ & - \frac{1}{32} E_g\{\ddot{A}(f_{2*}(X))^2(f(X) - g(X))^4\} \\ & \left. - \frac{1}{6} E_g\{\exp(t\Delta_Z(f, g))\Delta_Z(f, g)^3\} \right| \\ \leq & \left(\frac{1}{12} C_2^2 \varepsilon + \frac{1}{32} C_2^3 \varepsilon^2 + C_0^3 C_1 C_2 \exp(C_2 \varepsilon/4 + C_2/(4C_0)) \varepsilon \right) V(f - g) \\ < & V(f - g)/16, \end{aligned}$$

which implies $V(f - g)/16 \leq d_H^2(\mathbf{P}_f, \mathbf{P}_g) \leq 3V(f - g)/16$. This proves Lemma S.8. \square

Let ε satisfy the conditions in Lemma S.8. Define $\mathcal{F}_n = \{f \in S^m(\mathbb{I}) : \|f - f_0\|_\infty \leq \varepsilon/2, J(f) \leq (M + J(f_0)^{1/2})^2 r_n^2 \lambda^{-1}\}$. Note that for any $f \in \mathcal{F}_n$, we have $\|f\|_\infty \leq \|f_0\|_\infty + \varepsilon/2 < C$. Therefore, $\mathcal{F}_n \subseteq \mathcal{F}(C)$. Let $\mathcal{P}_n = \{\mathbf{P}_f^n : f \in \mathcal{F}_n\}$ and $D(\delta, \mathcal{P}_n, d_H)$ be the δ -packing number in terms of d_H . Since $r_n^2 \geq \lambda$ which leads to $(M + J(f_0)^{1/2}) r_n h^{-m} > M + J(f_0)^{1/2} > \varepsilon + \|f_0\|_\infty$, it can be easily checked that $\mathcal{F}_n \subset (M + J(f_0)^{1/2}) r_n h^{-m} \mathcal{T}$, where $\mathcal{T} = \{f \in S^m(\mathbb{I}) : \|f\|_\infty \leq 1, J(f) \leq 1\}$.

For any $f, g \in \mathcal{F}_n$ (implying $f, g \in \mathcal{F}(C)$) with $\|f - g\|_\infty \leq \varepsilon$, it follows by Lemma S.8 that $D(\delta, \mathcal{P}_n, d_H) \leq D(4\delta/\sqrt{3}, \mathcal{F}_n, d_V)$, where d_V is the distance induced by V , i.e., $d(f, g) = V^{1/2}(f - g)$. And hence, it follows by [29, Theorem 9.21] that

$$\begin{aligned} \log D(\delta, \mathcal{P}_n, d_H) &\leq \log D(4\delta/\sqrt{3}, \mathcal{F}_n, d_V) \\ &\leq \log D(4\delta/\sqrt{3}, (M + J(f_0)^{1/2}) r_n h^{-m} \mathcal{T}, d_V) \\ &\leq c_V \left(\frac{\delta}{(M + J(f_0)^{1/2}) r_n h^{-m}} \right)^{-1/m}, \end{aligned}$$

where c_V is a universal constant only depending on the regularity level m . This implies that for any $\delta > 2r_n$,

$$\begin{aligned} \log D(\delta/2, \mathcal{P}_n, d_H) &\leq \log D(r_n, \mathcal{P}_n, d_H) \\ &\leq c_V (M + J(f_0)^{1/2})^{1/m} h^{-1} \\ &\leq c_V (M + J(f_0)^{1/2})^{1/m} n r_n^2, \end{aligned}$$

where the last inequality follows by the fact $r_n^2 \geq (nh)^{-1}$. Thus, the right side of the above inequality is constant in δ . By [20, Theorem 7.1], with $\delta = Mr_n/4$, there exists test $\tilde{\phi}_n$ and a

universal constant $k_0 > 0$ satisfying

$$\begin{aligned}
E_{f_0}\{\tilde{\phi}_n\} &= \mathbf{P}_{f_0}^n \tilde{\phi}_n \\
&\leq \frac{\exp(c_V(M + J(f_0)^{1/2})^{1/m} n r_n^2) \exp(-k_0 n \delta^2)}{1 - \exp(-k_0 n \delta^2)} \\
&= \frac{\exp(c_V(M + J(f_0)^{1/2}) n r_n^2 - k_0 M^2 n r_n^2 / 16)}{1 - \exp(-k_0 M^2 n r_n^2 / 16)},
\end{aligned}$$

and, combined with Lemma S.8,

$$\begin{aligned}
\sup_{\substack{f \in \mathcal{F}_n \\ d_V(f, f_0) \geq 4\delta}} E_f\{1 - \tilde{\phi}_n\} &= \sup_{\substack{f \in \mathcal{F}_n \\ d_V(f, f_0) \geq 4\delta}} \mathbf{P}_f^n\{1 - \tilde{\phi}_n\} \\
&\leq \sup_{\substack{f \in \mathcal{F}_n \\ d_H(\mathbf{P}_f^n, \mathbf{P}_{f_0}^n) \geq \delta}} \mathbf{P}_f^n\{1 - \tilde{\phi}_n\} \\
&\leq \exp(-k_0 n \delta^2) = \exp(-k_0 M^2 n r_n^2 / 16).
\end{aligned}$$

This implies that

$$\begin{aligned}
&E_{f_0}\left\{\int_{\substack{f \in \mathcal{F}_n \\ d_V(f, f_0) \geq 4\delta}} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-n\lambda J(f)/2) d\Pi(f) (1 - \tilde{\phi}_n)\right\} \\
&\leq \int_{\substack{f \in \mathcal{F}_n \\ d_V(f, f_0) \geq 4\delta}} E_{f_0}\left\{\prod_{i=1}^n (p_f/p_{f_0})(Z_i) (1 - \tilde{\phi}_n)\right\} d\Pi(f) \\
&= \int_{\substack{f \in \mathcal{F}_n \\ d_V(f, f_0) \geq 4\delta}} E_f\{1 - \tilde{\phi}_n\} d\Pi(f) \\
&\leq \exp(-k_0 M^2 n r_n^2 / 16).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\int_{\substack{f \in \mathcal{F}_n \\ d_V(f, f_0) \geq 4\delta}} \prod_{i=1}^n (p_f/p_{f_0})(Z_i) \exp(-n\lambda J(f)/2) d\Pi(f) (1 - \tilde{\phi}_n) \\
&= O_{\mathbf{P}_{f_0}^n}(\exp(-k_0 M^2 n r_n^2 / 16)).
\end{aligned}$$

(S.23)

It follows from (S.20) and (S.22) that

$$P(A'_{n2} | \mathbf{D}_n) = O_{\mathbf{P}_{f_0}^n} \left(\exp(-(M - J(f_0)^{1/2})^2 n r_n^2 / 2 + c_2 n r_n^2) \right) = o_{\mathbf{P}_{f_0}^n}(1),$$

where the last inequality follows by $(M - J(f_0)^{1/2})^2 > 2(c_2 + 1)$ and $\exp(-n r_n^2) = o(1)$. Together

with (S.17), we get that

$$\begin{aligned}
& P(A'_n | \mathbf{D}_n) \\
& \leq P(A'_{n1} | \mathbf{D}_n) + P(A'_{n2} | \mathbf{D}_n) \\
& \leq P(A'_{n1}, \|f - f_0\|_\infty \leq \varepsilon/2 | \mathbf{D}_n) + P(\|f - f_0\|_\infty > \varepsilon/2 | \mathbf{D}_n) + P(A'_{n2} | \mathbf{D}_n) \\
& \leq P(A'_{n1}, \|f - f_0\|_\infty \leq \varepsilon/2 | \mathbf{D}_n) + o_{\mathbf{P}_{f_0}^n}(1) \\
& \leq P(A'_{n1}, \|f - f_0\|_\infty \leq \varepsilon/2 | \mathbf{D}_n) \tilde{\phi}_n \\
& \quad + P(A'_{n1}, \|f - f_0\|_\infty > \varepsilon/2 | \mathbf{D}_n) (1 - \tilde{\phi}_n) + o_{\mathbf{P}_{f_0}^n}(1).
\end{aligned}$$

Choose the constant M to be even bigger so that

$$c_V(M + J(f_0)^{1/2}) + 1 < k_0 M^2 / 16, \quad 1 + c_2 \leq k_0 M^2 / 16.$$

Then we get that

$$\begin{aligned}
& E_{f_0} \{P(A'_{n1}, \|f - f_0\|_\infty \leq \varepsilon/2 | \mathbf{D}_n) \tilde{\phi}_n\} \\
& \lesssim \exp(c_V(M + J(f_0)^{1/2})nr_n^2 - k_0 M^2 nr_n^2 / 16) \\
& \leq \exp(-nr_n^2) = o(1),
\end{aligned}$$

leading to $P(A'_{n1}, \|f - f_0\|_\infty \leq \varepsilon/2 | \mathbf{D}_n) \tilde{\phi}_n = o_{\mathbf{P}_{f_0}^n}(1)$. Meanwhile, it follows by (S.20) and (S.23) that

$$\begin{aligned}
& P(A'_{n1}, \|f - f_0\|_\infty \leq \varepsilon/2 | \mathbf{D}_n) (1 - \tilde{\phi}_n) \\
& \leq P(f \in \mathcal{F}_n, d_V(f, f_0) \geq 4\delta | \mathbf{D}_n) (1 - \tilde{\phi}_n) \\
& \leq \frac{\int_{d_V(f, f_0) \geq 4\delta} \prod_{i=1}^n (p_f / p_{f_0})(Z_i) \exp(-n\lambda J(f)/2) d\Pi(f) (1 - \tilde{\phi}_n)}{I_1} \\
& = O_{\mathbf{P}_{f_0}^n}(\exp(-k_0 M^2 nr_n^2 / 16 + c_2 nr_n^2)) \\
& = O_{\mathbf{P}_{f_0}^n}(\exp(-nr_n^2)) = o_{\mathbf{P}_{f_0}^n}(1).
\end{aligned}$$

Thus, we have shown that $P(\|f - f_0\| \geq \sqrt{2}Mr_n | \mathbf{D}_n) = o_{\mathbf{P}_{f_0}^n}(1)$. This completes the proof. \square

PROOF OF THEOREM 4.1. Fix any $\varepsilon_1, \varepsilon_2 \in (0, 1)$. Let $C = C_3 \sqrt{J(f_0)} + 1$, and C_0, C_1, C_2 be positive constants satisfying (2.2) and (2.3) in Assumptions A1. It follows by Lemma S.6 that for any fixed constant $M > 1$, if we set

$$(S.24) \quad b = \frac{C_2 C}{C_3} \sqrt{1 + \frac{1}{\rho_{m+1}}}, r = (nh / \log 2s)^{-1/2}, \delta_n = 2bh^m + 24C_0 c_K (4C_1 + M)r,$$

$$(S.25) \quad a_n = C_2 c_K^2 M^{1/2} h^{-1/2} r B(h) \delta_n, \text{ and } b_n = C_2^2 c_K h^{-1/2} \delta_n^2,$$

then as $n \rightarrow \infty$,

$$\mathbf{P}_{f_0}^n \left(\|\widehat{f}_{n,\lambda} - f_0\| \geq \delta_n \right) \leq 6n^{-M} \rightarrow 0,$$

and

$$\mathbf{P}_{f_0}^n \left(\|\widehat{f}_{n,\lambda} - f_0 - S_{n,\lambda}(f_0)\| > a_n + b_n \right) \leq 8n^{-M} \rightarrow 0.$$

By $B(h) \lesssim h^{-\frac{2m-1}{4m}}$ and the simple fact $a_n + b_n \lesssim D_n$, we get that

$$(S.26) \quad \|\widehat{f}_{n,\lambda} - f_0 - S_{n,\lambda}(f_0)\| = O_{\mathbf{P}_{f_0}^n}(a_n + b_n) = O_{\mathbf{P}_{f_0}^n}(D_n).$$

Recall that

$$S_{n,\lambda}(f_0) = \frac{1}{n} \sum_{i=1}^n (Y_i - \dot{A}(f_0(X_i))) K_{X_i} - \mathcal{P}_\lambda f_0.$$

It was shown by [41] that $\mathcal{P}_\lambda \varphi_\nu = \frac{\lambda \varphi_\nu}{1 + \lambda \varphi_\nu} \varphi_\nu$. Since f_0 satisfies Condition (S),

$$\begin{aligned} \|\mathcal{P}_\lambda f_0\|^2 &= \left\langle \sum_{\nu=1}^{\infty} f_\nu^0 \frac{\lambda \rho_\nu}{1 + \lambda \rho_\nu} \varphi_\nu, \sum_{\nu=1}^{\infty} f_\nu^0 \frac{\lambda \rho_\nu}{1 + \lambda \rho_\nu} \varphi_\nu \right\rangle \\ &= \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \frac{\lambda^2 \rho_\nu^2}{1 + \lambda \rho_\nu} \\ &= \lambda^{1 + \frac{\beta-1}{2m}} \sum_{\nu=1}^{\infty} |f_\nu^0|^2 \rho_\nu^{1 + \frac{\beta-1}{2m}} \frac{(\lambda \rho_\nu)^{1 - \frac{\beta-1}{2m}}}{1 + \lambda \rho_\nu} = O(h^{2m+\beta-1}), \end{aligned}$$

where the last equation follows by $\lambda = h^{2m}$, $\sup_{x \geq 0} \frac{x^{1 - \frac{\beta-1}{2m}}}{1+x} < \infty$, and Condition (S). On the other side, it follows by the proof of (S.11) that

$$\begin{aligned} &\mathbf{P}_{f_0}^n \left(\left\| \sum_{i=1}^n (Y_i - \dot{A}(f_0(X_i))) K_{X_i} \right\| \geq L(M) n(nh/\log 2)^{-1/2} \right) \\ &\leq 2 \exp(-Mnh(nh/\log 2)^{-1}) = 2^{1-M} \rightarrow 0, \text{ as } M \rightarrow \infty, \end{aligned}$$

implying that

$$\left\| \sum_{i=1}^n (Y_i - \dot{A}(f_0(X_i))) K_{X_i} \right\| = O_{\mathbf{P}_{f_0}^n}(n(nh/\log 2)^{-1/2}),$$

and hence,

$$\|S_{n,\lambda}(f_0)\| = O_{\mathbf{P}_{f_0}^n}((nh)^{-1/2} + h^{m + \frac{\beta-1}{2}}) = O_{\mathbf{P}_{f_0}^n}(\tilde{r}_n).$$

Together with (S.26) and the rate condition $D_n \lesssim \tilde{r}_n$, we get that

$$\|\widehat{f}_{n,\lambda} - f_0\| = O_{\mathbf{P}_{f_0}^n}(\tilde{r}_n).$$

Let M_1 be large constant so that event

$$(S.27) \quad \mathcal{E}'_n = \{\|\widehat{f}_{n,\lambda} - f_0\| \leq M_1 \tilde{r}_n\}$$

has probability approaching one. Meanwhile, from some positive constant M_0 , it follows by Theorem A.1 that $P(\|f - f_0\| \geq M_0 r_n | \mathbf{D}_n)$ converges to zero in $\mathbb{P}_{f_0}^n$ -probability. Let $C' > M_1$ be a constant to be further determined later, then we have that

$$\begin{aligned} & P(\|f - f_0\| \geq 2C' \tilde{r}_n | \mathbf{D}_n) \\ & \leq P(\|f - f_0\| \geq M_0 r_n | \mathbf{D}_n) + P(2C' \tilde{r}_n \leq \|f - f_0\| \leq M_0 r_n | \mathbf{D}_n). \end{aligned}$$

Thanks to Theorem A.1, the first term converges to zero in $\mathbb{P}_{f_0}^n$ -probability. Thus, when n is sufficiently large,

$$\mathbb{P}_{f_0}^n (P(\|f - f_0\| \geq M_0 r_n | \mathbf{D}_n) \geq \varepsilon_2/2) \leq \varepsilon_1/2.$$

We only need to handle the second term.

Define

$$(S.28) \quad \mathcal{E}_n'' = \left\{ \sup_{g \in \mathcal{G}} \|Z_{n,f_0}^{(l)}(g)\| \leq B(h) \sqrt{M \log n}, \quad l = 1, 2 \right\},$$

where

$$Z_{n,f_0}^{(l)}(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\psi_{n,f_0}^{(l)}(Z_i; g) K_{X_i} - E_f \{\psi_{n,f_0}^{(l)}(Z_i; g) K_{X_i}\}] \text{ for } l = 1, 2,$$

and

$$\begin{aligned} \psi_{n,f_0}^{(1)}(Z_i; g) &= c_K^{-1} h^{1/2} g(X_i), \\ \psi_{n,f_0}^{(2)}(Z_i; g) &= C_2^{-1} c_K^{-1} h^{1/2} \ddot{A}(f_0(X_i)) g(X_i). \end{aligned}$$

It is easy to see that $\psi_{n,f_0}^{(l)}(Z_i; g)$ satisfies (S.7). By Lemma S.5 we have that \mathcal{E}_n'' has $\mathbb{P}_{f_0}^n$ -probability approaching one. Thus, it holds that, when n becomes large, $\mathbb{P}_{f_0}^n(\mathcal{E}_n) \geq 1 - \varepsilon_1/2$. In the rest of the proof we simply assume that $\mathcal{E}_n \equiv \mathcal{E}_n' \cap \mathcal{E}_n''$ holds.

Let $I_n(f) = \int_0^1 \int_0^1 s D S_{n,\lambda}(\hat{f}_{n,\lambda} + s s'(f - \hat{f}_{n,\lambda}))(f - \hat{f}_{n,\lambda})(f - \hat{f}_{n,\lambda}) ds ds'$ for any $f \in S^m(\mathbb{I})$. Let $\Delta f = f - \hat{f}_{n,\lambda}$. Therefore,

$$\begin{aligned} & I_n(f) \\ &= -\frac{1}{n} \int_0^1 \int_0^1 s \sum_{i=1}^n \ddot{A}(\hat{f}_{n,\lambda}(X_i) + s s'(\Delta f)(X_i)) (\Delta f)(X_i)^2 ds ds' \\ & \quad - \lambda J(\Delta f, \Delta f)/2 \\ &= -\frac{1}{n} \int_0^1 \int_0^1 s \sum_{i=1}^n [\ddot{A}(\hat{f}_{n,\lambda}(X_i) + s s'(\Delta f)(X_i)) (\Delta f)(X_i)^2 \\ & \quad - \ddot{A}(f_0(X_i)) (\Delta f)(X_i)^2] ds ds' \\ & \quad - \frac{1}{2n} \sum_{i=1}^n [\ddot{A}(f_0(X_i)) (\Delta f)(X_i)^2 - E_{f_0}^X \{\ddot{A}(f_0(X)) (\Delta f)(X)^2\}] - \frac{1}{2} \|\Delta f\|^2 \\ &\equiv T_1(f) + T_2(f) - \frac{1}{2} \|\Delta f\|^2, \end{aligned}$$

where

$$\begin{aligned}
T_1(f) &= -\frac{1}{n} \int_0^1 \int_0^1 s \sum_{i=1}^n [\ddot{A}(\hat{f}_{n,\lambda}(X_i) + ss'(\Delta f)(X_i))(\Delta f)(X_i)^2 \\
&\quad - \ddot{A}(f_0(X_i))(\Delta f)(X_i)^2] ds ds', \\
T_2(f) &= -\frac{1}{2n} \sum_{i=1}^n [\ddot{A}(f_0(X_i))(\Delta f)(X_i)^2 - E_{f_0}^X \{\ddot{A}(f_0(X))(\Delta f)(X)^2\}].
\end{aligned}
\tag{S.29}$$

By Taylor's expansion in terms of Fréchet derivatives,

$$\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f}_{n,\lambda}) = S_{n,\lambda}(\hat{f}_{n,\lambda})(f - \hat{f}_{n,\lambda}) + I_n(f) = I_n(f).$$

Therefore,

$$\begin{aligned}
P(A_n | \mathbf{D}_n) &= \frac{\int_{A_n} \exp(n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f}_{n,\lambda}))) d\Pi(f)}{\int_{S^m(\mathbb{I})} \exp(n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f}_{n,\lambda}))) d\Pi(f)} \\
&= \frac{\int_{A_n} \exp(nI_n(f)) d\Pi(f)}{\int_{S^m(\mathbb{I})} \exp(nI_n(f)) d\Pi(f)},
\end{aligned}$$

where $A_n = \{f \in S^m(\mathbb{I}) : 2C'\tilde{r}_n \leq \|f - f_0\| \leq M_0 r_n\}$.

Let

$$J_1 = \int_{S^m(\mathbb{I})} \exp(nI_n(f)) d\Pi(f), \quad J_2 = \int_{A_n} \exp(nI_n(f)) d\Pi(f).$$

Then on \mathcal{E}_n and for $\|f - f_0\| \leq \tilde{r}_n$, we have $\|f - \hat{f}_{n,\lambda}\| \leq \|f - f_0\| + \|\hat{f}_{n,\lambda} - f_0\| \leq (M_1 + 1)\tilde{r}_n$.

Let $d_n = c_K(M_1 + 1)h^{-1/2}\tilde{r}_n$. It follows by similar arguments above (S.12) that $d_n^{-1}\Delta f \in \mathcal{G}$. It follows by Lemma S.4 that $\|\Delta f\|_\infty \leq c_K h^{-1/2}\|\Delta f\| \leq c_K(M_1 + 1)h^{-1/2}\tilde{r}_n$. By rate assumption $r_n = o(h^{3/2})$ and $h^{1/2} \log N = o(1)$ and the simple fact $\tilde{r}_n \leq r_n \sqrt{\log 2N}$, we get that

$$h^{-1/2}\tilde{r}_n \leq h^{-1/2}r_n \sqrt{\log 2N} = o(h\sqrt{\log N}) = o(1).$$

Therefore, we can let n be large so that, on \mathcal{E}_n and $\|f_0\|_\infty + \|\hat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty < C$. Then

on \mathcal{E}_n , we have

$$\begin{aligned}
& |T_1(f)| \\
& \leq C_2 \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2n} \sum_{i=1}^n (\Delta f)(X_i)^2 \\
& = C_2 \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2n} \sum_{i=1}^n [(\Delta f)(X_i)^2 - E^X\{(\Delta f)(X)^2\}] \\
& \quad + C_2 \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2} E^X\{(\Delta f)(X)^2\} \\
& \leq C_2 \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2n} \|\Delta f\| \\
& \quad \times \left\| \sum_{i=1}^n [(\Delta f)(X_i)K_{X_i} - E^X\{(\Delta f)(X)K_X\}] \right\| \\
& \quad + C_2 \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2} E^X\{(\Delta f)(X)^2\} \\
& \leq C_2 d_n \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2n} \|\Delta f\| \\
& \quad \times \left\| \sum_{i=1}^n [d_n^{-1}(\Delta f)(X_i)K_{X_i} - E^X\{d_n^{-1}(\Delta f)(X)K_X\}] \right\| \\
& \quad + C_2^2 \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2} \|\Delta f\|^2 \\
& \leq C_2 d_n \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2n} \|\Delta f\| \cdot c_K \sqrt{nh}^{-1/2} B(h) \sqrt{M \log N} \\
& \quad + C_2^2 \frac{\|\widehat{f}_{n,\lambda} - f_0\|_\infty + \|\Delta f\|_\infty}{2} \|\Delta f\|^2 \\
& \leq \frac{1}{2} C_2 M^{1/2} c_K^3 (2M_1 + 1)^3 h^{-3/2} \tilde{r}_n^3 n^{-1/2} B(h) \sqrt{\log N} \\
& \quad + \frac{1}{2} C_2^2 c_K (2M_1 + 1)^3 h^{-1/2} \tilde{r}_n^3 \\
& \leq D_1(C_2, c_K, M, M_1) \times \tilde{r}_n^3 (n^{-1/2} h^{-\frac{8m-1}{4m}} \sqrt{\log N} + h^{-1/2}) \\
& \leq D_1(C_2, c_K, M, M_1) \times \tilde{r}_n^3 b_{n1},
\end{aligned}$$

(S.30)

where $D_1(C_2, c_K, M, M_1)$ is constant depending only on C_2, c_K, M, M_1 .

We can use similar empirical processes techniques to handle T_2 . Note that on \mathcal{E}_n and for

$\|f - f_0\| \leq \tilde{r}_n$, using Assumption A1,

$$\begin{aligned}
& |T_2(f)| \\
&= \frac{1}{2n} \left| \sum_{i=1}^n [\ddot{A}(f_0(X_i))(\Delta f)(X_i)^2 - E_{f_0}^X \{\ddot{A}(f_0(X))(\Delta f)(X)^2\}] \right| \\
&= \frac{1}{2n} \left| \left\langle \sum_{i=1}^n [\ddot{A}(f_0(X_i))(\Delta f)(X_i)K_{X_i} - E_{f_0}^X \{\ddot{A}(f_0(X))(\Delta f)(X)K_X\}], \Delta f \right\rangle \right| \\
&\leq \frac{1}{2n} \|\Delta f\| \\
&\quad \times \left\| \sum_{i=1}^n [\ddot{A}(f_0(X_i))(\Delta f)(X_i)K_{X_i} - E_{f_0}^X \{\ddot{A}(f_0(X))(\Delta f)(X)K_X\}] \right\| \\
&= \frac{C_2 c_K h^{-1/2} d_n \|\Delta f\|}{2\sqrt{n}} \times \|Z_{n,f_0}^{(2)}(d_n^{-1} \Delta f)\| \\
&\leq \frac{C_2 c_K h^{-1/2} d_n \|\Delta f\|}{2\sqrt{n}} B(h) \sqrt{M \log N} \\
&\leq D_2(C_2, c_K, M, M_1) \times n^{-1/2} h^{-\frac{6m-1}{4m}} \tilde{r}_n^2 \sqrt{\log N} \\
&\leq D_2(C_2, c_K, M, M_1) \times \tilde{r}_n^2 b_{n2},
\end{aligned}
\tag{S.31}$$

where $D_2(C_2, c_K, M, M_1)$ is constant depending only on C_2, c_K, M_1, M .

It follows that on \mathcal{E}_n ,

$$\begin{aligned}
& J_1 \\
&\geq \int_{\|f - f_0\| \leq \tilde{r}_n} \exp(nI_n(f)) d\Pi(f) \\
&= \int_{\|f - f_0\| \leq \tilde{r}_n} \exp\left(nT_1(f) + nT_2(f) - \frac{n}{2} \|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f) \\
&\geq \exp(-[D_1(C_2, c_K, M, M_1) \tilde{r}_n b_{n1} + D_2(C_2, c_K, M, M_1) b_{n2} \\
&\quad + (M_1 + 1)^2/2] n \tilde{r}_n^2) \Pi(\|f - f_0\| \leq \tilde{r}_n).
\end{aligned}$$

To continue, we provide a lower bound for $\Pi(\|f - f_0\| \leq \tilde{r}_n)$ using the same arguments as in (S.19). Note that $\lambda \leq \tilde{r}_n^{\frac{4m}{2m+\beta-1}}$. Then it follows by Lemma S.7 (with d_n therein replaced by \tilde{r}_n)

and the proof of Theorem A.1 that

$$\begin{aligned}
& \Pi(\|f - f_0\| \leq \tilde{r}_n) \\
&= \mathbb{P}(\|G - f_0\| \leq \tilde{r}_n) \\
&\geq \mathbb{P}(V(G - f_0) \leq \tilde{r}_n^2/2, \lambda J(G - f_0) \leq \tilde{r}_n^2/2) \\
&\geq \mathbb{P}(V(G - f_0) \leq \tilde{r}_n^2/2, J(G - f_0) \leq \tilde{r}_n^{\frac{2(\beta-1)}{2m+\beta-1}}/2) \\
&= \mathbb{P}(\tilde{V}(\tilde{G} - \tilde{f}_0) \leq \tilde{r}_n^2/2, \tilde{J}(\tilde{G} - \tilde{f}_0) \leq \tilde{r}_n^{\frac{2(\beta-1)}{2m+\beta-1}}/2) \\
&\geq \mathbb{P}(\tilde{V}(\tilde{G} - \omega) \leq (1/\sqrt{2} - 1/2)^2 \tilde{r}_n^2, \tilde{J}(\tilde{G} - \omega) \leq (1/\sqrt{2} - 1/2)^2 \tilde{r}_n^{\frac{2(\beta-1)}{2m+\beta-1}}) \\
&\geq \exp(-\|\omega\|_\beta^2/2) \\
&\quad \times \mathbb{P}(\tilde{V}(\tilde{G}) \leq (1/\sqrt{2} - 1/2)^2 \tilde{r}_n^2, \tilde{J}(\tilde{G}) \leq (1/\sqrt{2} - 1/2)^2 \tilde{r}_n^{\frac{2(\beta-1)}{2m+\beta-1}}) \\
&\geq \exp(-\|\omega\|_\beta^2/2) \mathbb{P}(\tilde{V}(\tilde{G}) \leq (1/\sqrt{2} - 1/2)^2 \tilde{r}_n^2/2) \\
&\quad \times \mathbb{P}(\tilde{J}(\tilde{G}) \leq (1/\sqrt{2} - 1/2)^2 \tilde{r}_n^{\frac{2(\beta-1)}{2m+\beta-1}}/2) \\
&\geq \exp(-c_3 \tilde{r}_n^{-\frac{2}{2m+\beta-1}}),
\end{aligned}$$

where $c_3 > 0$ is a universal constant. Note that

$$\tilde{r}_n \geq (nh)^{-1/2} + h^{m+\frac{\beta-1}{2}} \geq 2n^{-\frac{2m+\beta-1}{2(2m+\beta)}},$$

we get that

$$n\tilde{r}_n^{2+\frac{2}{2m+\beta-1}} \geq n(4n^{-\frac{2m+\beta-1}{2m+\beta}})^{1+\frac{1}{2m+\beta-1}} = 4.$$

Therefore, $\tilde{r}_n^{-\frac{2}{2m+\beta-1}} \leq n\tilde{r}_n^2/4$, leading to

$$(S.32) \quad \Pi(\|f - f_0\| \leq \tilde{r}_n) \geq \exp\left(-\frac{c_3}{4} n\tilde{r}_n^2\right).$$

This implies by rate conditions $\tilde{r}_n b_{n1} \leq 1$ and $b_{n2} \leq 1$ that, on \mathcal{E}_n ,

$$\begin{aligned}
J_1 &\geq \exp(-[D_1(C_2, c_K, M, M_1)\tilde{r}_n b_{n1} + D_2(C_2, c_K, M, M_1)b_{n2} \\
&\quad + (M_1 + 1)^2/2 + c_3/4]n\tilde{r}_n^2) \\
&\geq \exp(-[D_1(C_2, c_K, M, M_1) + D_2(C_2, c_K, M, M_1) \\
&\quad + (M_1 + 1)^2/2 + c_3/4]n\tilde{r}_n^2).
\end{aligned}$$

Next we handle J_2 . The idea is similar to how we handle J_1 but with technical difference. Let $\Delta f = f - \hat{f}_{n,\lambda}$. Note that $\tilde{r}_n \leq r_n \sqrt{\log n}$, and hence, on \mathcal{E}_n , for any $f \in A_n$, i.e., $\|f - f_0\| \leq M_0 r_n$, we get that $\|\Delta f\| = \|\hat{f}_{n,\lambda} - f\| \leq \|\hat{f}_{n,\lambda} - f_0\| + \|f - f_0\| \leq M_1 \tilde{r}_n + M_0 r_n \leq (M_0 + M_1) r_n \sqrt{\log n}$. This implies that on \mathcal{E}_n , $\|\Delta f\|_\infty \leq c_K(M_0 + M_1) h^{-1/2} r_n \sqrt{\log n}$, where the last term by our rate assumption is $o(1)$, and hence, we can choose n to be large enough so that $\|f_0\|_\infty + \|\hat{f}_{n,\lambda} - f_0\|_\infty +$

$\|\Delta f\|_\infty < C$. Let $d_{*n} = c_K(M_0 + M_1)h^{-1/2}r_n\sqrt{\log n}$. Then $d_{*n}^{-1}\Delta f \in \mathcal{G}$. Using previous similar arguments handling $T_1(f)$, we have that on \mathcal{E}_n , for any $f \in A_n$,

$$\begin{aligned}
& |T_1(f)| \\
& \leq \frac{C_2 c_K (2M_1 + M_0)}{2n} h^{-1/2} r_n \sqrt{\log n} \\
& \quad \times \left(d_{*n} \left\| \sum_{i=1}^n [d_{*n}^{-1}(\Delta f)(X_i) K_{X_i} - E^X \{d_{*n}^{-1}(\Delta f)(X) K_X\}] \right\| \cdot \|\Delta f\| \right. \\
& \quad \left. + n E^X \{(\Delta f)(X)^2\} \right) \\
& \leq \frac{C_2 c_K (2M_1 + M_0)}{2n} h^{-1/2} r_n \sqrt{\log n} \\
& \quad \times (\sqrt{n} c_K h^{-1/2} d_{*n} \cdot (M_0 + M_1) r_n \sqrt{\log n} \cdot B(h) \sqrt{M \log N} \\
& \quad + n C_2 [(M_0 + M_1) r_n \sqrt{\log n}]^2) \\
& = \frac{1}{2} C_2 c_K^3 (2M_1 + M_0)^3 M^{1/2} h^{-3/2} r_n^3 n^{-1/2} B(h) (\log n)^2 \\
& \quad + \frac{1}{2} C_2^2 c_K (2M_1 + M_0)^3 h^{-1/2} r_n^3 (\log n)^{3/2} \\
& \leq D_3(C_2, c_K, M, M_0, M_1) \times r_n^3 \left(n^{-1/2} h^{-\frac{8m-1}{4m}} (\log n)^2 + h^{-1/2} (\log n)^{3/2} \right) \\
& = D_3(C_2, c_K, M, M_0, M_1) \times r_n^3 b_{n1} \leq D_3(C_2, c_K, M, M_0, M_1) \times \tilde{r}_n^2,
\end{aligned}$$

where $D_3(C_2, c_K, M, M_0, M_1)$ is constant depending only on C_2, c_K, M, M_0, M_1 and the last inequality follows by rate condition $r_n^3 b_{n1} \leq \tilde{r}_n^2$. Likewise, on \mathcal{E}_n and for any $f \in A_n$,

$$\begin{aligned}
|T_2(f)| & \leq \frac{\|\Delta f\|}{2\sqrt{n}} C_2 c_K h^{-1/2} d_{*n} \cdot B(h) \sqrt{M \log n} \\
& \leq \frac{1}{2} C_2 c_K^2 (M_0 + M_1)^2 M^{1/2} n^{-1/2} h^{-1} r_n^2 B(h) (\log n)^{3/2} \\
& \leq D_4(C_2, c_K, M, M_0, M_1) \times n^{-1/2} r_n^2 h^{-\frac{6m-1}{4m}} (\log n)^{3/2} \\
& = D_4(C_2, c_K, M, M_0, M_1) \times r_n^2 b_{n2} \leq D_4(C_2, c_K, M, M_0, M_1) \times \tilde{r}_n^2,
\end{aligned}$$

where $D_4(C_2, c_K, M, M_0, M_1)$ is constant only depending on C_2, c_K, M, M_0, M_1 and the last inequality follows by rate condition $r_n^2 b_{n2} \leq \tilde{r}_n^2$. It is easy to see that on \mathcal{E}_n and for any $f \in A_n$,

$$\|\hat{f}_{n,\lambda} - f\| \geq \|f - f_0\| - \|\hat{f}_{n,\lambda} - f_0\| \geq (2C' - M_1)\tilde{r}_n,$$

leading to that

$$J_2 \leq \exp \left(- \left(\frac{(2C' - M_1)^2}{2} - D_3(C_2, c_K, M, M_0, M_1) - D_4(C_2, c_K, M, M_0, M_1) \right) n \tilde{r}_n^2 \right).$$

Choose $C' > M_1$ to be large s.t.

$$\begin{aligned}
& \frac{(2C' - M_1)^2}{2} \geq \\
& 1 + D_1(C_2, c_K, M, M_1) + D_2(C_2, c_K, M, M_1) + D_3(C_2, c_K, M, M_0, M_1) \\
& + D_4(C_2, c_K, M, M_0, M_1) + (M_1 + 1)^2/2 + c_3/4.
\end{aligned}$$

Therefore, on \mathcal{E}_n ,

$$P(A_n|\mathbf{D}_n) \leq \frac{J_2}{J_1} \leq \exp(-n\tilde{r}_n^2).$$

When n becomes large s.t. $\exp(-n\tilde{r}_n^2) \leq \varepsilon_2/2$, we get that

$$\mathbf{P}_{f_0}^n(P(A_n|\mathbf{D}_n) \geq \varepsilon_2/2) \leq \mathbf{P}_{f_0}^n(\mathcal{E}_n^c) \leq \varepsilon_1/2.$$

This shows that

$$\mathbf{P}_{f_0}^n(P(\|f - f_0\| \geq 2C'\tilde{r}_n|\mathbf{D}_n) \geq \varepsilon_2) \leq \varepsilon_1.$$

Proof is completed. \square

S.4. Proof in Section 5.

PROOF OF THEOREM 5.1. Let $\varepsilon_1, \varepsilon_2$ be arbitrarily small positive constants. Let ε_3 be small fixed with $0 < \varepsilon_3 < \log 2$ and $4\varepsilon_3 \exp(\varepsilon_3) + 2\varepsilon_3 \leq \varepsilon_2/3$.

Choose $C = C_3\sqrt{J(f_0)} + 1$, and let C_0, C_1, C_2 be positives satisfying Assumption A1. For a constant $M > 1$, let b, r, δ_n, a_n, b_n be the same as in (S.24) and (S.25), and define $\mathcal{E}'_n = \{\|\hat{f}_{n,\lambda} - f_0\| \leq M_1\tilde{r}_n\}$ (with a large constant M_1) and \mathcal{E}''_n to be the same as in (S.28), such that with $\mathbf{P}_{f_0}^n$ -probability greater than $1 - \varepsilon_1/3$, both \mathcal{E}'_n and \mathcal{E}''_n hold. It follows from Theorem 4.1 that there exist a large constant $M_2 > 0$ s.t. for large n , $\mathbf{P}_{f_0}^n(\mathcal{E}'''_n) \geq 1 - \varepsilon_1/3$, where $\mathcal{E}'''_n = \{P(\|f - f_0\| \geq M_2\tilde{r}_n|\mathbf{D}_n) \leq \varepsilon_3\}$. Similar to the proof of Theorem 4.1, it can be shown that on $\mathcal{E}'_n \cap \mathcal{E}''_n$,

$$\begin{aligned} & P_0(\|f - f_0\| \geq M_2\tilde{r}_n) \\ &= \frac{\int_{\|f-f_0\| \geq M_2\tilde{r}_n} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f)}{\int_{S^m(\mathbb{I})} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f)} \\ &\leq \frac{\int_{\|f-f_0\| \geq M_2\tilde{r}_n} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f)}{\int_{\|f-f_0\| \geq \tilde{r}_n} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f)} \\ (S.33) \quad &\leq \exp\left(-((M_2 - M_1)^2/2 - (M_1 + 1)^2/2 - c_3/4)\right), \end{aligned}$$

where $c_3 > 0$ is the universal constant appearing in (S.32). We can even manage M_2 to be large so that the quantity (S.33) is less than $\exp(-n\tilde{r}_n^2)$. When n is large so that $\exp(-n\tilde{r}_n^2) \leq \varepsilon_3$, we get that $\mathbf{P}_{f_0}^n(\mathcal{E}'''_n) \geq \mathbf{P}_{f_0}^n(\mathcal{E}'_n \cap \mathcal{E}''_n) \geq 1 - \varepsilon_1/3$, where $\mathcal{E}'''_n = \{P_0(\|f - f_0\| \geq M_2\tilde{r}_n) \leq \varepsilon_3\}$. Define $\mathcal{E}_n = \mathcal{E}'_n \cap \mathcal{E}''_n \cap \mathcal{E}'''_n \cap \mathcal{E}''''_n$, then it can be seen that $\mathbf{P}_{f_0}^n(\mathcal{E}_n) \geq 1 - \varepsilon_1$.

Let T_1 and T_2 be defined in (S.29). Then

$$(S.34) \quad \ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f}_{n,\lambda}) + \frac{1}{2}\|f - \hat{f}_{n,\lambda}\|^2 = T_1(f) + T_2(f).$$

By (S.30) and (S.31) in the proof of Theorem 4.1, it can be shown that on \mathcal{E}_n , for any $f \in S^m(\mathbb{I})$ satisfying $\|f - f_0\| \leq M_2\tilde{r}_n$,

$$(S.35) \quad |T_1(f)| \leq D_5 \times \tilde{r}_n^3 b_{n1}, \quad |T_2(f)| \leq D_6 \times \tilde{r}_n^2 b_{n2},$$

where $D_5 = D_5(C_2, c_K, M, M_1, M_2)$ and $D_6 = D_6(C_2, c_K, M, M_1, M_2)$ are positive constants depending only on C_2, c_K, M, M_1, M_2 . Since $n\tilde{r}_n^2(\tilde{r}_n b_{n1} + b_{n2}) = o(1)$, we choose n to be large enough so that $D_5 \times n\tilde{r}_n^3 b_{n1} + D_6 \times n\tilde{r}_n^2 b_{n2} \leq \varepsilon_3$.

Define

$$\begin{aligned} J_{n1} &= \int_{S^m(\mathbb{I})} \exp\left(n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f}_{n,\lambda}))\right) d\Pi(f), \\ J_{n2} &= \int_{S^m(\mathbb{I})} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f), \\ \bar{J}_{n1} &= \int_{\|f-f_0\| \leq M_2 \tilde{r}_n} \exp\left(n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f}_{n,\lambda}))\right) d\Pi(f), \\ \bar{J}_{n2} &= \int_{\|f-f_0\| \leq M_2 \tilde{r}_n} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f). \end{aligned}$$

It is easy to see that on \mathcal{E}_n ,

$$0 \leq \frac{J_{n1} - \bar{J}_{n1}}{J_{n1}} \leq \varepsilon_3, \quad 0 \leq \frac{J_{n2} - \bar{J}_{n2}}{J_{n2}} \leq \varepsilon_3.$$

By some algebra, it can be shown that the above inequalities lead to

$$(S.36) \quad (1 - \varepsilon_3) \cdot \frac{\bar{J}_{n2}}{\bar{J}_{n1}} \leq \frac{J_{n2}}{J_{n1}} \leq \frac{1}{1 - \varepsilon_3} \cdot \frac{\bar{J}_{n2}}{\bar{J}_{n1}}.$$

Meanwhile, on \mathcal{E}_n , using (S.35) and the elementary inequality $|\exp(x) - 1| \leq 2|x|$ for $|x| \leq \log 2$, we get that

$$\begin{aligned} & |\bar{J}_{n2} - \bar{J}_{n1}| \\ & \leq \int_{\|f-f_0\| \leq M_2 \tilde{r}_n} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) \\ & \quad \times |\exp(n(T_1(f) + T_2(f))) - 1| d\Pi(f) \\ & \leq 2\varepsilon_3 \bar{J}_{n2}, \end{aligned}$$

leading to that

$$(S.37) \quad \frac{1}{1 + 2\varepsilon_3} \leq \frac{\bar{J}_{n2}}{\bar{J}_{n1}} \leq \frac{1}{1 - 2\varepsilon_3}.$$

Combining (S.36) and (S.37), on \mathcal{E}_n ,

$$\frac{1 - \varepsilon_3}{1 + 2\varepsilon_3} \leq \frac{J_{n2}}{J_{n1}} \leq \frac{1}{(1 - 2\varepsilon_3)(1 - \varepsilon_3)},$$

leading to

$$(S.38) \quad -4\varepsilon_3 \leq \frac{1 - \varepsilon_3}{1 + 2\varepsilon_3} - 1 \leq \frac{J_{n2}}{J_{n1}} - 1 \leq \frac{1}{(1 - 2\varepsilon_3)(1 - \varepsilon_3)} - 1 \leq 4\varepsilon_3$$

For simplicity, denote $R_n(f) = n(T_1(f) + T_2(f))$. For any $S \in \mathcal{S}$, let $S' = S \cap \{f \in S^m(\mathbb{I}) : \|f - f_0\| \leq M_2 \tilde{r}_n\}$. Then on \mathcal{E}_n , we get that

$$|P(S|\mathbf{D}_n) - P_0(S)| \leq |P(S'|\mathbf{D}_n) - P_0(S')| + 2\varepsilon_3.$$

Moreover, it follows from (S.38) that on \mathcal{E}_n ,

$$\begin{aligned} & |P(S'|\mathbf{D}_n) - P_0(S')| \\ = & \left| \int_{S'} \left(\frac{\exp(n(\ell_{n,\lambda}(f) - \ell_{n,\lambda}(\hat{f}_{n,\lambda})))}{J_{n1}} - \frac{\exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right)}{J_{n2}} \right) d\Pi(f) \right| \\ \leq & \int_{S'} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) \times \left| \frac{\exp(R_n(f))}{J_{n1}} - \frac{1}{J_{n2}} \right| d\Pi(f) \\ \leq & \int_{S'} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) \times \frac{|\exp(R_n(f)) - 1|}{J_{n2}} d\Pi(f) \\ & + \int_{S'} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) \times \exp(R_n(f)) \times \left| \frac{1}{J_{n1}} - \frac{1}{J_{n2}} \right| d\Pi(f) \\ \leq & 2\varepsilon_3 \frac{\int_{S'} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f)}{J_{n2}} \\ & + \exp(\varepsilon_3) \times \left| \frac{1}{J_{n1}} - \frac{1}{J_{n2}} \right| \times \int_{S'} \exp\left(-\frac{n}{2}\|f - \hat{f}_{n,\lambda}\|^2\right) d\Pi(f) \\ \leq & 2\varepsilon_3 + \exp(\varepsilon_3) \times \left| \frac{J_{n2}}{J_{n1}} - 1 \right| \leq 2\varepsilon_3 + 4\varepsilon_3 \exp(\varepsilon_3) \leq \varepsilon_2/3. \end{aligned}$$

Note that the right hand side is free of S . Then we get that on \mathcal{E}_n ,

$$\sup_{S \in \mathcal{S}} |P(S|\mathbf{D}_n) - P_0(S)| \leq \varepsilon_2/3 + 2\varepsilon_3 \leq \varepsilon_2.$$

This implies that for sufficiently large n ,

$$\begin{aligned} & \mathbb{P}_{f_0}^n \left(\sup_{S \in \mathcal{S}} |P(S|\mathbf{D}_n) - P_0(S)| > \varepsilon_2 \right) \\ \leq & \mathbb{P}_{f_0}^n(\mathcal{E}_n^c) + \mathbb{P}_{f_0}^n \left(\mathcal{E}_n, \sup_{S \in \mathcal{S}} |P(S|\mathbf{D}_n) - P_0(S)| > \varepsilon_2 \right) \\ = & \mathbb{P}_{f_0}^n(\mathcal{E}_n^c) \leq \varepsilon_1. \end{aligned}$$

This completes the proof. \square

S.5. L^2 -diameter of $R_n^\omega(\alpha)$. Without additional restrictions, the L^2 -diameter of $R_n^\omega(\alpha)$ in (6.2) is infinity. To see this, consider $f = \tilde{f}_{n,\lambda} + \sum_{\nu=1}^N f_\nu \varphi_\nu$, where $f_\nu^2 = \frac{r_{\omega,n}(\alpha)^2}{N\omega_\nu}$ for $1 \leq \nu \leq N$. Then $f \in R_n^\omega(\alpha)$ since $\sum_{\nu=1}^N \omega_\nu f_\nu^2 = r_{\omega,n}(\alpha)^2$. However,

$$\|f\|_2^2 = \sum_{\nu=1}^N f_\nu^2 = \frac{r_{\omega,n}(\alpha)^2}{N} \sum_{\nu=1}^N \omega_\nu^{-1} \geq \frac{r_{\omega,n}(\alpha)^2}{N} \sum_{\nu=1}^N \nu = \frac{r_{\omega,n}(\alpha)^2(N+1)}{2}.$$

Letting $N \rightarrow \infty$, we can see that $\|f\|_2^2 \rightarrow \infty$. Therefore, the L^2 -diameter of $R_n^\omega(\alpha)$ is infinity.

Next we investigate the L^2 -diameter of $R_n^{\star\omega}(\alpha)$. For any $g, f \in R_n^{\star\omega}(\alpha)$, let $u = g - f \equiv \sum_{\nu=1}^{\infty} u_\nu \varphi_\nu$, and choose $J_n \sim n^{1/(2m+\beta)}$. It follows by Remark 4.1 in the revised manuscript that $r_{\omega,n}(\alpha) = O_{P_{f_0}}(n^{-1/2})$, and hence, $\|u\|_\omega \leq 2r_{\omega,n}(\alpha) = O_{P_{f_0}}(n^{-1/2})$. Then

$$\begin{aligned} \|u\|_2^2 &= \sum_{1 \leq \nu \leq J_n} u_\nu^2 + \sum_{\nu > J_n} u_\nu^2 \\ &= \sum_{1 \leq \nu \leq J_n} \omega_\nu u_\nu^2 \omega^{-1} + \sum_{\nu \geq J_n} \rho_\nu^{1+\frac{\beta-1}{2m}} u_\nu^2 \rho_\nu^{-(1+\frac{\beta-1}{2m})} \\ &\leq 4J_n \log(2J_n) r_{\omega,n}(\alpha)^2 + 4M J_n^{-(2m+\beta-1)} \\ &= O_{P_{f_0}}(n^{-\frac{2m+\beta-1}{2m+\beta}} \log n), \end{aligned}$$

indicating that the L^2 -diameter of $R_n^{\star\omega}(\alpha)$ is $O_{P_{f_0}}(n^{-\frac{2m+\beta-1}{2(2m+\beta)}} \sqrt{\log n})$.

S.6. Proof of Proposition 6.5.

PROOF OF PROPOSITION 6.5. Under the setup of Proposition 6.5, it follows from [41] that $\ddot{A}(\cdot) \equiv 1$, and hence, (2.4) becomes the following uniform free beam problem:

$$(S.39) \quad \varphi_\nu^{(4)}(\cdot) = \rho_\nu \varphi_\nu(\cdot), \quad \varphi_\nu^{(j)}(0) = \varphi_\nu^{(j)}(1) = 0, \quad j = 2, 3.$$

The eigenvalues satisfy $\lim_{\nu \rightarrow \infty} \rho_\nu / (\pi\nu)^4 = 1$; see [24, Problem 3.10]. The normalized solutions to (S.39) are

$$(S.40) \quad \varphi_1(z) = 1, \quad \varphi_2(z) = \sqrt{3}(2z - 1),$$

$$(S.41) \quad \varphi_{2k+1}(z) = \frac{\sin(\gamma_{2k+1}(z - 1/2))}{\sin(\gamma_{2k+1}/2)} + \frac{\sinh(\gamma_{2k+1}(z - 1/2))}{\sinh(\gamma_{2k+1}/2)}, \quad k \geq 1.$$

$$(S.42) \quad \varphi_{2k+2}(z) = \frac{\cos(\gamma_{2k+2}(z - 1/2))}{\cos(\gamma_{2k+2}/2)} + \frac{\cosh(\gamma_{2k+2}(z - 1/2))}{\cosh(\gamma_{2k+2}/2)}, \quad k \geq 1,$$

where $\gamma_\nu = \rho_\nu^{1/4}$ satisfying $\cos(\gamma_\nu) \cosh(\gamma_\nu) = 1$; see [5, page 295–296].

Proof of (i). By direct examinations, it can be shown that when $\nu \geq 3$ is odd, $\cos(x) \cosh(x) = 1$ has a unique solution in $((\nu + 1/2)\pi, (\nu + 1)\pi)$, that is, $\gamma_\nu \in ((\nu + 1/2)\pi, (\nu + 1)\pi)$; when $\nu \geq 3$ is even, $\cos(x) \cosh(x) = 1$ has a unique solution in $(\nu\pi, (\nu + 1/2)\pi)$, that is, $\gamma_\nu \in (\nu\pi, (\nu + 1/2)\pi)$. Consequently, for any $k \geq 1$, $0 < \gamma_{2k+2} - \gamma_{2k+1} < \pi$.

Let δ_0 be constant such that $0 < \delta_0 < \pi/2 - \pi|z - 1/2|$, and $d_0 = \min\{\sin^2(\delta_0), \cos^2(\delta_0 + \pi|z - 1/2|)\}$. Clearly, $d_0 > 0$ is a constant. It is easy to see that when $k \rightarrow \infty$,

$$\frac{\sinh(\gamma_{2k+1}(z - 1/2))}{\sinh(\gamma_{2k+1}/2)} \rightarrow 0, \quad \text{and} \quad \frac{\cosh(\gamma_{2k+2}(z - 1/2))}{\cosh(\gamma_{2k+2}/2)} \rightarrow 0.$$

Then for arbitrarily small $\varepsilon \in (0, d_0/8)$, there exists N s.t. for any $k \geq N$,

$$\varphi_{2k+1}(z)^2 \geq \frac{1}{2} \sin^2(\gamma_{2k+1}(z - 1/2)) - \varepsilon, \text{ and}$$

$$\varphi_{2k+2}(z)^2 \geq \frac{1}{2} \cos^2(\gamma_{2k+2}(z - 1/2)) - \varepsilon.$$

Let $\phi'_k = (\gamma_{2k+2} - \gamma_{2k+1})(z - 1/2)$. Then $|\phi'_k| \leq \pi|z - 1/2| < \pi/2$. There exists an integer l_k s.t. $\gamma_{2k+1}(z - 1/2) = \phi_k + l_k\pi$, where $\phi_k \in [0, \pi)$. Then

$$\sin^2(\gamma_{2k+1}(z - 1/2)) = \sin^2(\phi_k), \text{ and } \cos^2(\gamma_{2k+2}(z - 1/2)) = \cos^2(\phi_k + \phi'_k).$$

If $0 \leq \phi_k \leq \delta_0$, then it can be seen that

$$-\pi|z - 1/2| \leq \phi'_k \leq \phi_k + \phi'_k \leq \delta_0 + \phi'_k \leq \delta_0 + \pi|z - 1/2|.$$

Therefore, $\cos^2(\phi_k + \phi'_k) \geq \cos^2(\delta_0 + \pi|z - 1/2|)$. If $\delta_0 < \phi_k < \pi - \delta_0$, then $\sin^2(\phi_k) \geq \sin^2(\delta_0)$. If $\pi - \delta \leq \phi_k < \pi$, then it can be seen that

$$\pi - \delta_0 - \pi|z - 1/2| \leq \phi_k + \phi'_k < \pi + \pi|z - 1/2|.$$

Therefore, $\cos^2(\phi_k + \phi'_k) \geq \cos^2(\delta_0 + \pi|z - 1/2|)$. Consequently, for any $k \geq N$,

$$\begin{aligned} & \varphi_{2k+1}(z)^2 + \varphi_{2k+2}(z)^2 \\ & \geq \frac{1}{2} (\sin^2(\gamma_{2k+1}(z - 1/2)) + \cos^2(\gamma_{2k+2}(z - 1/2))) - 2\varepsilon \\ & \geq \frac{1}{2} \min\{\sin^2(\delta_0), \cos^2(\delta_0 + \pi|z - 1/2|)\} - 2\varepsilon \geq d_0/4. \end{aligned}$$

Then we have

$$\begin{aligned} & \sum_{\nu > 2} \frac{h\varphi_\nu(z)^2}{(1 + \lambda\rho_\nu + (\lambda\rho_\nu)^{1+\beta/4})^j} \\ & = \sum_{k \geq 1} \frac{h\varphi_{2k+1}(z)^2}{(1 + \lambda\rho_{2k+1} + (\lambda\rho_{2k+1})^{1+\beta/4})^j} + \sum_{k \geq 1} \frac{h\varphi_{2k+2}(z)^2}{(1 + \lambda\rho_{2k+2} + (\lambda\rho_{2k+2})^{1+\beta/4})^j} \\ & \geq \sum_{k \geq 1} \frac{h\varphi_{2k+1}(z)^2 + h\varphi_{2k+2}(z)^2}{(1 + \lambda\rho_{2k+2} + (\lambda\rho_{2k+2})^{1+\beta/4})^j} \\ & \geq \sum_{k \geq N} \frac{hd_0/4}{(1 + \lambda\rho_{2k+2} + (\lambda\rho_{2k+2})^{1+\beta/4})^j} \\ & \gtrsim \sum_{k \geq N} \frac{h}{(1 + (k\pi h)^4 + (k\pi h)^{4+\beta})^j} \\ & \geq \int_N^\infty \frac{h}{(1 + (\pi hx)^4 + (\pi hx)^{4+\beta})^j} dx \\ & = \frac{1}{\pi} \int_{\pi Nh}^\infty \frac{1}{(1 + x^4 + x^{4+\beta})^j} dx \xrightarrow{h \rightarrow 0} \frac{1}{\pi} \int_0^\infty \frac{1}{(1 + x^4 + x^{4+\beta})^j} dx > 0. \end{aligned}$$

This shows that condition (6.5) holds for $r = 1$.

Proof of (ii). Write $\omega = \sum_{\nu} \omega_{\nu} \varphi_{\nu}$ where ω_{ν} is a square-summable real sequence. Then $F_{\omega}(\varphi_{\nu}) = \int_0^1 \omega(z) \varphi_{\nu}(z) dz = \omega_{\nu}$. Therefore, $\sum_{\nu} F_{\omega}(\varphi_{\nu})^2 = \sum_{\nu} \omega_{\nu}^2 < \infty$. Meanwhile, since $\omega \neq 0$, $\sum_{\nu=1}^{\infty} F_{\omega}(\varphi_{\nu})^2 > 0$. Consequently, for $j = 1, 2$, it follows by dominated convergence theorem that as $n \rightarrow \infty$,

$$\sum_{\nu=1}^m \frac{F_{\omega}(\varphi_{\nu})^2}{(1 + \lambda + n^{-1} \sigma_{\nu}^{-2})^j} + \sum_{\nu > m} \frac{F_{\omega}(\varphi_{\nu})^2}{(1 + \lambda \rho_{\nu} + (\lambda \rho_{\nu})^{1+\beta/(2m)})^j} \rightarrow \sum_{\nu=1}^{\infty} F_{\omega}(\varphi_{\nu})^2 > 0.$$

Hence (6.5) holds for $r = 0$. □

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
250 N. UNIVERSITY STREET
WEST LAFAYETTE, IN 47906
EMAIL: SHANG9@PURDUE.EDU; CHENGG@STAT.PURDUE.EDU